

Pending Application 09/587,436

Filed on 06/05/00

Inventors: Havre *et al.*

ATTORNEY DOCKET NUMBER: 7991-086-99
SERIAL NUMBER: 09/685,403
REFERENCE: **AY**

REC2 KINASE

1. FIELD OF THE INVENTION

The present invention concerns the field of molecular genetics and medicine. Particularly, it concerns a gene encoding a protein that is a kinase and is involved in cell cycle regulation and the repair of damaged genomic DNA in mammalian cells. The gene and protein, termed herein, respectively *hsREC2* and *hsRec2*, is in the same supergene family as the mammalian protein having homologous pairing and strand transfer activities, *RAD51* and was isolated because of its homology to the homologous pairing and strand transfer protein of *Ustilago maydis*. Due to this relationship the same gene and protein is termed elsewhere *RAD51B* and *Rad51B*.

2. BACKGROUND OF THE INVENTION

2.1 THE STRUCTURE AND FUNCTION OF *hsREC2*

During the life of every organism the DNA of its cells is constantly subjected to chemical and physical events that cause alterations in its structure, i.e., potential mutations. These potential mutations are recognized by DNA repair enzymes found in the cell because of the mismatch between the strands of the DNA. To prevent the deleterious effects that would occur if these potential mutations became fixed, all organisms have a variety of mechanisms to repair DNA mismatches. In addition, higher animals have evolved mechanisms whereby cells having highly damaged DNA, undergo a process of programmed death ("apoptosis").

The association between defects in the DNA mismatch repair and apoptosis inducing pathways and the development, progression and response to treatment of oncologic disease is widely recognized, if incompletely understood, by medical scientists. Chung, D.C. & Rustgi, A.K., 1995, *Gastroenterology* 109:1685-99; Lowe, S.W., et al., 1994, *Science* 266:807-10. Therefore, there is a continuing need to identify and clone the genes that encode proteins involved in DNA repair and DNA mismatch monitoring.

Studies with bacteria, fungi and yeast have identified three genetically defined groups of genes involved in mismatch repair processes. The groups are termed,

respectively, the excision repair group, the error prone repair group and the recombination repair group. Mutants in a gene of each group result in a characteristic phenotype. Mutants in the recombination repair group in yeast result in a phenotype having extreme sensitivity to ionizing radiation, a sporulation deficiency, and decreased or absent mitotic recombination. Petes, T.D., et al., 1991, in Broach, J.R., et al., eds., THE MOLECULAR BIOLOGY OF THE YEAST *SACCHAROMYCES*, pp. 407-522 (Cold Spring Harbor Press, 1991).

Several phylogenetically related genes have been identified in the recombination repair group: *recA*, in *E. Coli*, Radding, C.M., 1989, Biochim. Biophys. Acta 1008:131-145; *RAD51* in *S. cerevisiae*, Shinohara, A., 1992, Cell 69:457-470, Aboussekhra, A.R., et al., 1992, Mol. Cell. Biol. 12:3224-3234, Basile, G., et al., 1992, Mol. Cell. Biol. 12:3235-3246; *RAD57* in *S. cerevisiae*, Gene 105:139-140; *REC2* in *U. maydis*, Bauchwitz, R., & Holloman, W.K., 1990, Gene 96:285-288, Rubin, B.P., et al., 1994, Mol. Cell. Biol. 14:6287-6296. A third *S. cerevisiae* gene *DMC1*, is related to *recA*, although mutants of *DMC1* show defects in cell-cycle progression, recombination and meiosis, but not in recombination repair.

The phenotype of *REC2* defective *U. maydis* mutants is characterized by extreme sensitivity to ionizing radiation, defective mitotic recombination and interplasmid recombination, and an inability to complete meiosis. Holliday, R., 1967, Mutational Research 4:275-288. UmREC2, the *REC2* gene product of *U. maydis*, has been extensively studied. It is a 781 amino acid ATPase that, in the presence of ATP, catalyzes the pairing of homologous DNA strands in a wide variety of circumstances, e.g., UmREC2 catalyzes the formation of duplex DNA from denatured strands, strand exchange between duplex and single stranded homologous DNA and the formation of a nuclease resistant complex between identical strands. Kmiec, E.B., et al., 1994, Mol. Cell. Biol. 14:7163-7172. UmREC2 is unique in that it is the only eukaryotic ATPase that forms homolog pairs, an activity it shares with the *E. coli* enzyme *recA*.

U.S. patent application, Serial No. 08/373,134, filed January 17, 1995, by W.K. Holloman and E.B. Kmiec discloses *REC2* from *U. maydis*, methods of producing recombinant UmREC2 and methods of its use. Prior to the date of the present invention a fragment of human *REC2* cDNA was available from the IMAGE consortium, Lawrence

Livermore National Laboratories, as plasmid p153195. Approximately 400 bp of the sequence of p153195 had been made publicly available on dbEST database.

The scientific publication entitled: ISOLATION OF HUMAN AND MOUSE GENES BASED ON HOMOLGY TO REC2, July 1997, Proc. Natl. Acad. Sci. **94**, 7417-7422 by Michael C. Rice et al., discloses the sequences of murine and human Rec2, of the human REC2 cDNA, and discloses that irradiation increases the level of hsREC2 transcripts in primary human foreskin fibroblasts. The scientific publication Albala et al., December 1997, Genomics **46**, 476-479 also discloses the sequence of the same protein and cDNA which it terms RAD51B. A sequence that is identical to hsREC2 except for the C-terminal 14 nucleotides of the coding sequence and the 3'-untranslated sequence was published by Cartwright R., et al., 1998, Nucleic Acids Research **26**, 1653-1659 and termed hsR51h2. It is believed that hsREC2 and hsR51h2 represent alternative processing of the same primary transcript. The parent application of this application was published as WO 98/11214 on March 19, 1998.

The structure of hsREC2 is also disclosed in application Serial No. 60/025,929, filed September 11, 1996, application Serial No. 08/927,165, filed September 11, 1997, and patent publication WO 98/11214, published March 19, 1998.

2.2 CELL CYCLE REGULATION

The eukaryotic cell cycle consists of four stages, G₁, S (synthesis), G₂, and M (mitosis). The underlying biochemical events that determine the stage of the cell and the rate of progression to the next stage is a series of kinases, e.g., cdk2, cdc2, which are regulated and activated by labile proteins that bind them, termed cyclins, e.g., cyclin D, cyclin E, Cyclin A. The activated complex in turn phosphorylates other proteins which activates the enzymes that are appropriate for each given stage of the cycle. Reviewed, Morgan, D.O., 1997, Ann. Rev. Cell. Dev. Biol. **15**, 261-291; Clurman, B.E., & Roberts, J.M., 1998, in THE GENETIC BASIS OF HUMAN CANCER, pp.173-191 (ed. by Vogelstein, B., & Kinzer K.W., McGraw Hill, NY) (hereafter Vogelstein)

The cell cycle contains a check point in G₁. Under certain conditions, e.g., chromosomal damage or mitogen deprivation, a normal cell will not progress beyond the check point. Rb and p53 are proteins involved in the G₁ check point related to mitogen

deprivation and chromosomal damage, respectively. Inactivating mutations in either of these proteins results, in concert with other mutations, in a growth transformed, i.e., malignant, cell. The introduction of a copy of the normal p53 or Rb gene suppresses the transformed phenotype. Accordingly genes, such as p53 or Rb, whose absence is associated with transformation are termed "tumor suppressor" genes. A frequent cause of familial neoplastic syndromes is the inheritance of a defective copy of a tumor suppressor gene. Reviewed Fearson, E.R., in *Vogelstein* pp. 229-236.

The level of p53 increases in response to chromosomal damage, however, the mechanism which mediates this response is unknown. It is known that p53 can be phosphorylated by a variety of kinases and that such phosphorylation may stabilize the p53 protein. Reviewed Agarwal, M.L., et al., Jan. 2, 1998, *J. Biol. Chem.* 273, 1-4.

3. SUMMARY OF THE INVENTION

The present invention is based on the unexpected discovery that hsRec2 is a serine kinase that phosphorylates several proteins that control the cell cycle, particularly cyclin E and p53. The invention permits the phosphorylation of the cell cycle control proteins at sites that are physiologically relevant. In addition, the discovery of the enzyme activity of Rec2 permits the construction of assays for the discovery of compounds that are specific antagonists and agonists of Rec2, which compounds have a pharmacological activity.

4. BRIEF DESCRIPTION OF THE FIGURES

Figures 1A-1D. Figures 1A and 1B show the derived amino acid sequence of hsREC2 (SEQ ID NO:1) and Figures 1C and 1D show the nucleic acid sequences of the *hsREC2* cDNA coding strand (SEQ ID NO:2). Figures 1E and 1F show the derived amino acid sequence of muREC2 (SEQ ID NO:3) and Figure 1G shows the nucleic acid sequences of the *muREC2* cDNA coding strand (SEQ ID NO:4).

Figure 2A-2C. Figure 2A is an annotated amino acid sequence of hsREC2. Specifically noted are the nuclear localization sequence ("NLS"), A Box and B Box motif sequences, DNA binding sequence and a src-type phosphorylation site ("P"). Figure 2B is a cartoon of the annotated sequence, showing in particular that the region 80-200 is most

closely related to recA. Figures 2C and 2D show the sequence homology between hsREC2 and *Ustilago maydis* REC2. The region of greatest similarity, 43% homology, is in bold.

Figure 3A-3B. A. The incorporation of ^{32}P -ATP into myelin basic protein (0.25 μM) as a function of time, concentration of Rec2 was 1 $\mu\text{g}/30\text{--}40\mu\text{l}$. B. The incorporation of ^{32}P -ATP into kemptide (LRRASLG, SEQ ID No: 5) during a 60 min. reaction as a function of kemptide concentration.

5. DETAILED DESCRIPTION OF THE INVENTION

As used herein, genes are all capitlized , e.g., hsREC2, while the corresponding protein is in initial capitalization, e.g., hsRec2.

The activity of hsREC2 was determined using an N-terminal hexahistadyl containing derivative that was produced in baculovirus. Confirming results were obtained with baculovirus produced glutathione-S-transferase conjugated hsREC2 and with thioredoxin-conjugated hsREC2 produced in *E. coli*. These confirming results tend to exclude that the kinase activity resulted from the co-purification of an endogenous baculovirus kinase on the Ni-NTA resin. To further exclude the possibility of purification artifacts the Ni-NTA purified hexahistadyl-hsREC2 was further purified by preparative SDS-PAGE. Only the fractions containing hsREC2 by silver stain were found to contain kinase activity.

The sequence of muRec2 and hsRec2 differ at only 56 of the 350 amino acids. The invention can be practiced using either muRec2 or hsRec2 or a protein that consists of a mixture of amino acids, i.e., at some positions the amino acid is that of muRec2 and at others the amino acid is that of hsRec2, hereafter a chimeric hs/muRec2. In addition, the mutein having a substitution for the tyrosine at position 163 can be used to practice the invention, e.g., Tyr-Ala . Thus, the invention can be further practiced using a chimeric hs/muREC2^{ala163}. In one embodiment the substitution can be any aliphatic amino acid. In an alternative embodiment the substitution can be any amino acid other than cysteine or proline. The term "Rec2 kinase" is used herein to denote the genus consisting of hsRec2, muRec2 and all chimeric hs/muRec2 proteins and the Tyr¹⁶³ substituted derivatives of each. The term artificial Rec2 kinase is a Rec2 kinase that is not

also a mammalian Rec2. The term mammalian Rec2 is used herein to denote the genus of proteins consisting of the mammalian homologs of hsRec2 and of muRec2.

The invention can further be practiced using a fusion protein, which consists of a protein having a sequence that comprises that of a Rec2 kinase or a mammalian Rec2 that is fused to a second sequence which is a protein or peptide that can be used to purify the resultant fusion protein.

The naturally occurring hsRec2 and muRec2 are found as phosphoproteins, the phosphorylation of which is not essential to the activity of the proteins as a kinase. In, the invention the terms Rec2 kinase and mammalian Rec2 encompass both the phosphorylated and non-phosphorylated forms of the proteins.

5.1 Cell Cycle Regulation

An expression vector comprising hsREC2 operably linked to the CMV immediate early promoter was constructed and transfected into CHO cells. A mutein was constructed in which tyrosine-163, a phosphorylatable tyrosine in an *src* site (phe-pro-arg-tyr) (amino acids 8-11 of SEQ ID No. 8) was replaced by alanine (hsREC2^{ala163}). Sham (neo^r) transfected, hsREC2 transfected and hsREC2^{ala163} transfected CHO cells were synchronized by serum starvation, released, and the DNA content was assayed by quantitative fluorescent flow cytometry at various time points. The hsREC2 transfected cells showed delayed onset of S phase. Thus, at 14 hours post release 75% of the hsREC2 transfected cells were in G₁ compared to 36% of the controls.

Over expression of hsREC2 but not hsREC2^{ala163} sensitizes the cell to UV radiation. CHO cells were irradiated with UV at 15 J/m². Again the cells were analyzed by quantitative fluorescent flow cytometry. The hsREC2 cells showed extensive apoptosis compared to the controls at 24, 48 and 72 hours post irradiation.

5.2 Kinase Activity

The kinase activity of hsREC2 can be shown on a variety of substrates. Artifactual substrates such as myelin basic protein, which is a known substrate for protein kinase C and protein kinase A are phosphorylated by hsREC2. The kemptide (leu-arg-arg-ala-ser-leu-gly), which is also a known substrate of ser/thr kinases can be phosphorylated. In

addition the following recombinantly produced proteins are phosphorylated by hsREC2: p53, cyclin B1 and cyclin E. The heterodimers of cyclin B1/cdc2 and cyclin E/cdk2 are also phosphorylated by hsREC2. The interpretation of these experiments is complicated by the fact that cyclin E/cdk2 autophosphorylates and that cyclin B1/cdc2 but not cyclin E/cdk2 phosphorylates hsREC2 itself. In contrast to the cyclinB1/cdc2 complex, hsRec2 is not an autophosphorylase.

Although expression of hsREC2^{ala163} in a cell has no effect on the cell cycle, the hsREC2^{ala163} mutein has full kinase activity.

Compounds having pharmacological activity with respect to mREC2 can be identified by assaying the kinase activity of an mREC2, and particularly hsREC2, in the presence of candidate agonists or antagonists. The particular preferred substrates are cyclin E and p53.

5.3 hsREC2 Association With Other Proteins

An S³⁵-radiolabeled preparation of hsREC2 was made by coupled transcription translation in a reticulocyte lysate system. The preparation was mixed with an extract from HCT116 cells. In separate reactions monoclonal antibodies to various cell proteins were added and the antibody bound material isolated with Protein A Sepharose. The bound material was then analyzed by SDS-PAGE, and autoradiographed. The immunoprecipitate contained hsREC2 when anti-p53, anti-PCNA and anti-cdc2 monoclonals were used. No hsREC2 was precipitated when anti-cdc4 or anti-cdk4 monoclonals were employed.

5.4 An hsREC2 Agonist or Antagonist Has a Pharmacologic Activity

The activities of hsREC2 indicate that the modulation of its activity can sensitize or desensitize a cell to enter apoptosis as a result of incurring genetic damage, as for example by UV radiation, and can also protect or deprotect a cell from DNA damage by extending or shortening the G₁ and S periods. Agonist and antagonists of hsREC2 are compounds having activities of the type that medical practitioners desire. The discovery of compounds that are hsREC2 agonists or antagonists will be important in pharmaceutical science.

In one embodiment, the invention is a method of determining whether a given compound has such a pharmacological activity by measuring the effects of the compound on the kinase activity of hsREC2. In specific embodiments, the invention is a method wherein the relative effects of the compound on hsREC2 and on a second kinase are assessed. For example, a compound that is an agonist of hsREC2, but has little or no effect on cyclin D/cdk4 and cyclin E/cdk2 would cause cells to arrest in G₁ and undergo apoptosis in response to genetic damage. In particular embodiments, the kinase assay is done with a substrate that is selected from the group consisting of p53, cdc2, cdk2 or cyclin E. Alternatively, the substrate can be a model substrate such as myelin basic protein or kemptide (leu-arg-arg-ala-ser-leu-gly).

6. EXAMPLES

6.1 The production of recombinant hsREC2 protein by baculovirus infection of *Autographica californica*

To facilitate the construction of an *hsREC2* expression vector, restriction sites for XhoI and KpnI were appended by PCR amplification to a the *hsREC2* cDNA. The *hsREC2* cDNA starting at nt 71 was amplified using the forward primer 5'-GAG CTCGAG GGTACC C ATG GGT AGC AAG AAA C-3' (SEQ ID NO:6), which placed the XhoI and KpnI sites (underlined) 5' of the start codon. The recombinant molecule containing the entire coding sequence of *hsREC2* cDNA, can be removed using either XhoI or KpnI and the unique XbaI site located between nt 1270 and 1280 of SEQ ID NO:2.

A vector, pBacGSTSV, for the expression of HsREC2 in baculovirus infected *Spodoptera frugiperda* (Sf-9) insect cells (ATCC cell line No. CRL1711, Rockville MD), was obtained from Dr. Zailin Yu (Baculovirus Expression Laboratory, Thomas Jefferson University). The vector pVLGS was constructed by the insertion of a fragment encoding a *Schistosoma japonicum* glutathione S-transferase polypeptide and a thrombin cleavage site from pGEX-2T (described in Smith & Johnson, GENE 67:31 (1988)), which is hereby incorporated by reference, into the vector pVL1393. A polyA termination signal sequence was inserted into pVLGS to yield pBacGSTSV. A plasmid containing the 1.2 Kb *hsREC2* fragment was cut with KpnI, the 3' unpaired ends removed with T4 polymerase and the product cut with XbaI. The resultant fragment was inserted into a SmaI, XbaI cut

pBacGSTSV vector to yield pGST/hsREC2.

Recombinant virus containing the insert from pGST/hsREC2 were isolated in the usual way and Sf-9 cells were infected. Sf-9 cells are grown in SF900II SFM (Gibco/BRL Cat # 10902) or TNM-FH (Gibco/BRL Cat # 11605-011) plus 10% FBS. After between 3-5 days of culture the infected cells are collected, washed in Ca^{++} and Mg^{++} free PBS and sonicated in 5ml of PBS plus proteinase inhibitors (ICN Cat # 158837), 1% NP-40, 250 mM NaCl per 5×10^7 cells. The lysate is cleared by centrifugation at 30,000 xg for 20 minutes. The supernatant is then applied to 0.5 ml of glutathione-agarose resin (Sigma Chem. Co. Cat # G4510) per 5×10^7 cells. The resin is washed in a buffer of 50 mM Tris-HCl, pH 8.0, 150 mM NaCl and 2.5 mM CaCl_2 , and the hsREC2 released by treatment with thrombin (Sigma Chem. Co. Cat # T7513) for 2 hours at 23°C in the same buffer. For certain experiments the thrombin is removed by the technique of Thompson and Davie, 1971, *Biochim Biophys Acta* 250:210, using an aminocaproyl-p-chlorobenzylamide affinity column (Sigma Chem. Co. Cat # A9527).

Alternatively, the full length hsREC2 cDNA was cloned into the expression vector, pAcHisA, for overexpression in a baculovirus system and purification utilizing a 6 histidine tag. For cloning, the hsREC2 expression cassette was cut with KpnI, the 3' protruding termini were removed with T4 polymerase, and the DNA was then digested with XbaI. The resulting fragment was ligated to pAcHisA using the SmaI and XbaI sites. Recombinant virus containing hsREC2 was purified and insect cells were infected by Dr. Z. Yu in the Baculovirus expression laboratory of the Kimmel Cancer Institute. Insect cell pellets from 2 liters of culture were suspended in 60 ml of 10 mM TrisCl, pH 7.5, 130 mM NaCl, 2% TX100, 2 $\mu\text{g/ml}$ leupeptin and aprotinin and 1 $\mu\text{g/ml}$ pepstatin and sonicated on ice 4 times for 5 seconds each using a microtip at a 20% pulse (Branson sonifier 450). Debris was removed by centrifuging at 30,000 xg for 20 minutes. The clarified supernatant was divided between two 50 ml culture tubes and 1 ml of Ni-NTA agarose added to each tube for 1 hour with rocking at 4°C. The unbound fraction was separated from the resin by a brief centrifugation and the resin was washed with 10 ml of 100 mM imidazole for 10 minutes on a rocker and centrifuged at 2000 rpm for 5 minutes. After a second 10 minute wash with 500 mM imidazole the slurry was transferred to a column and the effluent discarded. The purified his-hsRec2 was eluted

with 1M imidazole, pH 7.0 (imidazole on column for 10 minutes before collection of eluate), and dialyzed overnight against 50 mM TrisCl, pH 7.4, 50 mM NaCl, 10% glycerol. For simplicity, this protein will be referred to as hsRec2 instead of hishsRec2.

6.2 The Bacterial Production of recombinant hsREC2 protein

The hsREC2 cDNA coding region was excised from the previously used mammalian expression vector pcDNA3 G8 by cleavage with XbaI, removal of 5' protruding termini with T4 polymerase, followed by cleavage with KpnI. The resulting fragment was ligated into the KpnI and blunted HindIII sites of a bacterial expression vector pBAD/HisC (Invitrogen, Corp., USA). The constructed expression vector with hREC2 cloned in frame with a hexahistidine tag was electrotransformed into LMG194 bacteria (Invitrogen, Corp., USA) for expression. A 500ml LB ampicillin culture was inoculated by a single colony and grown at 37° into log phase. The culture was induced by .02% arabinose for 4 hours and harvested by centrifuging at 8,000 xg. The pellet was resuspended and lysed by 1mg/ml lysozyme and sonication in 5 volumes of 50mM NaH₂PO₄, 300mM NaCl, 1% TX100, 2µg/ml leupeptin and aprotinin and 1µg/ml pepstatin, .1 mg/ml DNase I, 10mM βME and 20mM imidazole at 0°C. The lysate was clarified by centrifugation at 10,000 xg for 30 minutes then added to a sealed column containing 1 ml activated Ni + NTA agarose resin and rocked at 4 for 1 hour. The column was then opened and washed by gravity with 20 volumes of 50mM NaH₂PO₄, 300mM NaCl, 1% TX100, 50mM imidazole at 4°. The bound protein was then eluted in 3 volumes of the above wash buffer with 500mM imidazole and collected in 1ml fractions. The purified His-HsRec2 was dialyzed over night against 50mM Tris, 50mM NaCl, 10% glycerol and stored at -80°.

6.3 Detection of hsREC2 Kinase

Phosphokinase filter assays. Substrates were either kemptide or myelin basic protein and approximately 1 µg of his-hsRec2 was added as the phosphokinase. For both assays, the buffer contained 50 mM TrisCl, pH 7.5, 10 mM MgCl₂, 1 mM DTT. The second substrate, ³²P-ATP was constant at 50 µM with a specific activity of 1972 cpm/pmole (kemptide) and 2980 cpm/pmole (MBP). ³²P-ATP was added to initiate the reaction which was carried out at 30° C. for the indicated time. At the end of the reaction, 20 µl was spotted on phosphocellulose discs, washed twice with 10 ml per disc in 1% phosphoric acid and twice

in distilled water. Filters were counted in a Wallac Scintillation counter. Substrate without hsRec2 added was used as a control and counts were subtracted to obtain a zero point.

Myelin basic protein (0.25 μ M) was phosphorylated for between 0 and 25 minutes at the above conditions. Phosphate incorporation was linear with time and reached 1.2 pmole at 25 minutes. Kemptide from 0 to 0.15 mM was phosphorylated for 60 minutes. The rate of phosphate incorporation was linear with substrate concentration up to 0.06 mM, where a rate of 0.09 pmoles/minute was observed.

Two different hsRec2 conjugates, GST-hsRec2 and thioredoxin-hsRec2, also exhibited phosphokinase activity. Further evidence that this activity was not a contaminant, was obtained by immunoprecipitating hsREC2 using hybridoma supernatants, followed by assay for phosphokinase activity using p53 as a substrate as described below. These experiment confirmed that the kinase activity was precipitable by anti-hsREC2 monoclonal antibodies.

Two substrates that were not phosphorylated by hsRec2, were a tyrosine kinase substrate peptide containing one tyrosine, derived from the sequence surrounding the phosphorylation site in pp60^{src} (RRLIEDAEYAARG) (SEQ ID No. 7), and an hsRec2 peptide, residues 153-172 (VEIAESRFPRYFNIEEKLLL) (SEQ ID No. 8).

p53 phosphorylation. Human recombinant p53 (0.5 μ g, Pharmingen, San Diego, CA) was incubated with or without hsRec2 in 50 mM TrisCl, pH 7.4, 10 mM MgCl₂, and 1 mM DTT at 30° C. The reaction was initiated by the addition of ³² P-ATP (25 μ M ATP, 40 cpm/femtomole). At the end of each time point an equal volume of 2X loading buffer (5) was added and tubes were placed on ice until all tubes were collected. Samples were then heated at 100° C for 10 minutes and 13 μ l were run on Ready Gels (Bio-Rad Laboratories, Hercules, CA), and transferred to nitrocellulose overnight prior to exposure to X-ray film. Radiolabeled p53 was readily observed.

cdc2/cyclin B phosphokinase assay. Purified human recombinant cyclin B1/cdc2 (Oncogene, Cambridge, MA), was incubated with hsRec2 for 10 or 60 minutes at 30° C., using the same buffer conditions as described for p53. An equal volume of 2X gel lading buffer was added (5), samples were heated at 100° C for 10 minutes and run on an SDS gel, transferred to nitrocellulose and exposed to film. Radiolabeled cyclin B1 due to hsREC2 kinase activity was readily observed above the level of "autophosphorylation" of cyclin B1 by cdc2. Radiolabeled cdc2 was observed only in the hsREC2 containing reactions mixtures

at 60 minutes but not at 10 minutes reaction time.

cdk2/cyclin E phosphokinase assay. GST-cyclin E was isolated from *E. coli* transformed with pGEX-2TcycE (A. Giordano, Thomas Jefferson University) and purified using Glutathione Sepharose 4B (Pharmacia, Piscataway, NJ). The glutathione Sepharose GST-cyclin E was washed, and then stored as a 1:1 slurry in 50 mM Tris Cl, pH 7.4. For assays with cyclin E bound cdk2, purified cdk2 (kindly given to us by A. Koff, Sloan-Kettering, NY) was incubated with cyclin E as described (6) and unbound cdk2 removed by washing prior to storage as a 1:1 slurry. Kinase assays were carried out with the immobilized GST-cyclin E with or without bound cdk2 otherwise using the same conditions described for p53. Phosphorylation of cyclin E and hsREC2 was readily observed in the absence of cdk2. In the presence of cdk2, autophosphorylation was seen, however, hsREC2 phosphorylation of cyclin E above that level was readily apparent.

***In vitro* associated between p53 and hsRec2.** HsRec2 (5 μ g) and 15 μ l agarose-GST-p53 (Oncogene Sciences) were added to 0.5 ml of binding buffer (10% glycerol, 50 mM Tris Cl, pH 7.4, 0.1 mM EDTA, 1mM DTT, 0.02% NP40, 200mM NaCl, 10 μ g/ml aprotinin and leupeptin, and 20 μ M PMSF. Following one hour at room temperature, the p53 agarose was pelleted and washed twice with buffer as above, using a higher concentration of detergent (0.1% NP40), and once with 50mM TrisCl, pH 7.4, 10mM MgCl₂.

Association of *in vitro* translated hsRec2 with PCNA, p53 and cdc2. XbaI linearized pCMVhREC2 was first transcribed *in vitro* (Ambion, Austin TX) using 1 μ g of the vector, and then translated *in vitro* along with Xef1 mRNA included in the kit as a positive control. Reticulocyte lysates containing Xef1 or hsRec2 translation products labeled with ³⁵S-methionine were incubated with 1.2 mg cell extract from HCT116 cells (50 mM TrisCl, pH 7.4, 120 mM NaCl, 0.5% NP40, 20 μ M PMSF, 2 μ g/ml pepstatin, and 10 μ g/ml leupeptin and aprotinin, MB) for 2 hours, then 10 μ g of antibodies against PCNA, p53 or cdc2 were added for an overnight incubation. On the following day, Protein A Sepharose was added for 2 hours, and pellets were washed four times with 500 μ l MB. Pellets were suspended in 40 μ l of sample buffer, boiled 10 minutes and 15 μ l run on a 10% gel, then transferred to nitrocellulose to obtain a lower background, before exposure to X-ray film.

CLAIMS:

1. A method of phosphorylating a serine-containing substrate which comprises incubating the substrate with an effective concentration of ATP and an enzyme having a sequence which comprises the sequence of a Rec2 kinase or a mammalian Rec2 and measuring the amount of phosphorylation of the substrate.
2. The method of claim 1, wherein the sequence of the enzyme comprises the sequence of a Rec2 kinase containing other than a Tyr¹⁶³.
3. The method of claim 2, wherein the sequence of the enzyme comprises the sequence of hsRec2 containing other than a Tyr¹⁶³.
4. The method of claim 3, wherein the substrate is selected from the group consisting of the human proteins p53, cdc2, cdk2 and cyclin E.
5. The method of claim 3, wherein the substrate is a kemptide.
6. The method of claim 1, wherein the sequence of the enzyme comprises the sequence of hsRec2.
7. The method of claim 6, wherein the substrate is selected from the group consisting of p53, cdc2, cdk2 or cyclin E.
8. The method of claim 6, wherein the substrate is a kemptide.
9. The method of claim 1, wherein the sequence of the enzyme comprises the sequence of a mammalian Rec2.
10. The method of claim 9, wherein the substrate is selected from the group consisting of

the human proteins p53, cdc2, cdk2 and cyclin E.

11. The method of claim 9, wherein the substrate is a kemptide.
12. The method of claim 1, which further comprises the steps of forming a mixture of the enzyme and an antagonist or an agonist of the enzyme and measuring the effect of said antagonist or agonist on the amount of phosphorylation on the substrate.
13. A composition comprising
 - a. an enzyme having a sequence that comprises the sequence of a Rec2 kinase or a mammalian Rec2;
 - b. a serine-containing substrate of the enzyme; and
 - c. a γ -phosphate labeled ATP.
14. The composition of claim 13, in which the labeled phosphate is a ^{32}P .
15. The composition of claim 13, in which the substrate is a cell-cycle control protein.
16. The composition of claim 15 in which the substrate is a protein selected from the group consisting of human p53, human cdc2, human cdk2 and human cyclin E.
17. The composition of claim 13, in which the substrate is a kemptide.
18. The composition of claim 13, in which the sequence of the enzyme comprises the sequence of hsRec2 or hsRec2^{Ala163}.
19. An enzyme comprising a Rec2 kinase having an amino acid that is other than a Tyr¹⁶³.
20. An enzyme having a sequence comprising the sequence of a mammalian Rec2 having an amino acid that is other than a Tyr¹⁶³.

ABSTRACT

The invention includes a method of phosphorylating a serine containing substrate by incubating the substrate with ATP and an enzyme that is hsRec2 or muRec2 or a derivative thereof. The natural substrates of the kinase activity of Rec2 are the cell cycle control proteins such as p53 and cyclin E. The over expression of Rec2 is known to cause cell-cycle arrest and apoptosis and the invention discloses that these effects are kinase mediated. Accordingly, the invention provides a method of assessing antagonists and agonists of Rec2, which antagonists and agonists would have pharmacological activity. The invention further discloses that there is specific binding between hsRec2 and at least three cell cycle control proteins: p53, PCNA and cdc2.

SEQUENCE LISTING

<110> Havre, Pamela A.
Rice, Michael C.
Holloman, William K.
Kmiec, Eric B.

<120> REC2 Kinase

<130> 7991-034-999

<160> 8

<170> FastSEQ for Windows Version 3.0

<210> 1

<211> 350

<212> PRT

<213> Homo Sapiens

<400> 1

```

Met Gly Ser Lys Lys Leu Lys Arg Val Gly Leu Ser Gln Glu Leu Cys
 1           5           10          15
Asp Arg Leu Ser Arg His Gln Ile Leu Thr Cys Gln Asp Phe Leu Cys
      20           25          30
Leu Ser Pro Leu Glu Leu Met Lys Val Thr Gly Leu Ser Tyr Arg Gly
    35           40          45
Val His Glu Leu Leu Cys Met Val Ser Arg Ala Cys Ala Pro Lys Met
    50           55          60
Gln Thr Ala Tyr Gly Ile Lys Ala Gln Arg Ser Ala Asp Phe Ser Pro
   65           70          75          80
Ala Phe Leu Ser Thr Thr Leu Ser Ala Leu Asp Glu Ala Leu His Gly
      85           90          95
Gly Val Ala Cys Gly Ser Leu Thr Glu Ile Thr Gly Pro Pro Gly Cys
      100          105         110
Gly Lys Thr Gln Phe Cys Ile Met Met Ser Ile Leu Ala Thr Leu Pro
     115          120         125
Thr Asn Met Gly Gly Leu Glu Gly Ala Val Val Tyr Ile Asp Thr Glu
     130          135         140
Ser Ala Phe Ser Ala Glu Arg Leu Val Glu Ile Ala Glu Ser Arg Phe
    145          150         155         160
Pro Arg Tyr Phe Asn Thr Glu Glu Lys Leu Leu Leu Thr Ser Ser Lys
      165          170         175
Val His Leu Tyr Arg Glu Leu Thr Cys Asp Glu Val Leu Gln Arg Ile
      180          185         190
Glu Ser Leu Glu Glu Glu Ile Ile Ser Lys Gly Ile Lys Leu Val Ile
     195          200         205
Leu Asp Ser Val Ala Ser Val Val Arg Lys Glu Phe Asp Ala Gln Leu
     210          215         220
Gln Gly Asn Leu Lys Glu Arg Asn Lys Phe Leu Ala Arg Glu Ala Ser
    225          230         235         240
Ser Leu Lys Tyr Leu Ala Glu Glu Phe Ser Ile Pro Val Ile Leu Thr

```


<400> 3

```

Met Ser Ser Lys Lys Leu Arg Arg Val Gly Leu Ser Pro Glu Leu Cys
 1           5           10           15
Asp Arg Leu Ser Arg Tyr Leu Ile Val Asn Cys Gln His Phe Leu Ser
      20           25           30
Leu Ser Pro Leu Glu Leu Met Lys Val Thr Gly Leu Ser Tyr Arg Gly
      35           40           45
Val His Glu Leu Leu His Thr Val Ser Lys Ala Cys Ala Pro Gln Met
      50           55           60
Gln Thr Ala Tyr Glu Leu Lys Thr Arg Arg Ser Ala His Leu Ser Pro
      65           70           75           80
Ala Phe Leu Ser Thr Thr Leu Cys Ala Leu Asp Glu Ala Leu His Gly
      85           90           95
Gly Val Pro Cys Gly Ser Leu Thr Glu Ile Thr Gly Pro Pro Gly Cys
      100          105          110
Gly Lys Thr Gln Phe Cys Ile Met Met Ser Val Leu Ala Thr Leu Pro
      115          120          125
Thr Ser Leu Gly Gly Leu Glu Gly Ala Val Val Tyr Ile Asp Thr Glu
      130          135          140
Ser Ala Phe Thr Ala Glu Arg Leu Val Glu Ile Ala Glu Ser Arg Phe
      145          150          155          160
Pro Gln Tyr Phe Asn Thr Glu Glu Lys Leu Leu Leu Thr Ser Ser Arg
      165          170          175
Val His Leu Cys Arg Glu Leu Thr Cys Glu Gly Leu Leu Gln Arg Leu
      180          185          190
Glu Ser Leu Glu Glu Glu Ile Ile Ser Lys Gly Val Lys Leu Val Ile
      195          200          205
Val Asp Ser Ile Ala Ser Val Val Arg Lys Glu Phe Asp Pro Lys Leu
      210          215          220
Gln Gly Asn Ile Lys Glu Arg Asn Lys Phe Leu Gly Lys Gly Ala Ser
      225          230          235          240
Leu Leu Lys Tyr Leu Ala Gly Glu Phe Ser Ile Pro Val Ile Leu Thr
      245          250          255
Asn Gln Ile Thr Thr His Leu Ser Gly Ala Leu Pro Ser Gln Ala Asp
      260          265          270
Leu Val Ser Pro Ala Asp Asp Leu Ser Leu Ser Glu Gly Thr Ser Gly
      275          280          285
Ser Ser Cys Leu Val Ala Ala Leu Gly Asn Thr Trp Gly His Cys Val
      290          295          300
Asn Thr Arg Leu Ile Leu Gln Tyr Leu Asp Ser Glu Arg Arg Gln Ile
      305          310          315          320
Leu Ile Ala Lys Ser Pro Leu Ala Ala Phe Thr Ser Phe Val Tyr Thr
      325          330          335
Ile Lys Gly Glu Gly Leu Val Leu Gln Gly His Glu Arg Pro
      340          345          350

```

<210> 4

<211> 1525

<212> DNA

<213> Mus Musculus

<400> 4

```

gggagccctg gaaacatgag cagcaagaaa ctaagacgag tgggtttatc tccagagctg      60
tgtgaccgtt taagcagata cctgattggt aactgtcagc actttttaag tctctcccca      120
ctagaactta tgaaagtgcac tggcctgagt tacagaggtg tccacgagct tcttcataca      180

```

gtaagcaagg	cctgtgcccc	gcagatgcaa	acggcttatg	agttaaagac	acgaagggtct	240
gcacatctct	caccggcatt	cctgtctact	accctgtgcg	ccttgatga	agcattgcac	300
gggtgtgtgc	cttgtggatc	tctcacagag	attacaggtc	caccaggttg	cggaaaaact	360
cagttttgca	taatgatgag	tgtcttagct	acattaccta	ccagcctggg	aggattagaa	420
ggggctgtgg	tctacatcga	cacagagtct	gcatttactg	ctgagagact	ggttgagatt	480
gcggaatctc	gttttccaca	atattttaac	actgaggaaa	aattgcttct	gaccagcagt	540
agagttcatc	tttgccgaga	gctcacctgt	gaggggcttc	tacaaaggct	tgagtctttg	600
gaggaagaga	tcatttcgaa	aggagttaag	cttgtgattg	ttgactccat	tgcttctgtg	660
gtcagaaagg	agtttgacct	gaagcttcaa	ggcaacatca	aagaaaggaa	caagttcttg	720
ggcaaaaggag	cgtccttact	gaagtacctg	gcaggggagt	tttcaatccc	agttatcttg	780
acgaatcaaa	ttacgaccca	tctgagtgga	gccctccctt	ctcaagcaga	cctgggtgtct	840
ccagctgatg	atttgtccct	gtctgaaggc	acttctggat	ccagctgttt	ggtagctgca	900
ctaggaaca	catgggggtca	ctgtgtgaac	acccggctga	ttctccagta	ccttgattca	960
gagagaaggc	agattctcat	tgccaagtct	cctctggctg	ccttcacctc	ctttgtctac	1020
accatcaagg	gggaaggcct	ggttcttcaa	ggccacgaaa	gaccataggg	atactgtgac	1080
ctttgtctag	tgctgattgc	atgtgactca	tgaatgaaa	caggactgcg	ctgcttgga	1140
aaagaaacg	gaagccaaca	taatgaggat	taattgggtg	gttgctgttg	agggtgtaac	1200
agtgatttca	gacccggaag	gtgaagatga	agaagccttt	atccagtctc	tgtagtcaga	1260
ggctaggggc	tccaccaccg	tgggatgtca	gcggccatcg	taataatttg	cacttacaca	1320
agcacccttc	agccatgccc	ctcaaagtgg	ttcagccaca	ttaattaatt	aaagcccaca	1380
atccccctag	ggagagcagg	agggggacta	acaagatttg	taattacaga	agggaaaatt	1440
tccgaataaa	gtattgttcc	gccaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	1500
aaaaaaaaaa	aaaaaaaaaa	aaaaa				1525

<210> 5

<211> 7

<212> PRT

<213> Artificial Sequence

<220>

<223> Substrate of ser/thr kinases

<400> 5

Leu Arg Arg Ala Ser Leu Gly

1

5

<210> 6

<211> 32

<212> DNA

<213> Artificial Sequence

<220>

<223> PCR Primer

<400> 6

gagctcgagg gtacccatgg gtagcaagaa ac

32

<210> 7

<211> 13

<212> PRT

<213> Artificial Sequence

<220>

<223> Fragment of Naturally Occurring Protein

<400> 7
 Arg Arg Leu Ile Glu Asp Ala Glu Tyr Ala Ala Arg Gly
 1 5 10

<210> 8
 <211> 20
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Fragment of Naturally Occurring Protein

<400> 8
 Val Glu Ile Ala Glu Ser Arg Phe Pro Arg Tyr Phe Asn Thr Glu Glu
 1 5 10 15
 Lys Leu Leu Leu
 20

Met Gly Ser Lys Lys Leu Lys Arg Val Gly Leu Ser Gln Glu Leu Cys	1	5	10	15
Asp Arg Leu Ser Arg His Gln Ile Leu Thr Cys Gln Asp Phe Leu Cys	20	25	30	
Leu Ser Pro Leu Glu Leu Met Lys Val Thr Gly Leu Ser Tyr Arg Gly	35	40	45	
Val His Glu Leu Leu Cys Met Val Ser Arg Ala Cys Ala Pro Lys Met	50	55	60	
Gln Thr Ala Tyr Gly Ile Lys Ala Gln Arg Ser Ala Asp Phe Ser Pro	65	70	75	80
Ala Phe Leu Ser Thr Thr Leu Ser Ala Leu Asp Glu Ala Leu His Gly	85	90	95	
Gly Val Ala Cys Gly Ser Leu Thr Glu Ile Thr Gly Pro Pro Gly Cys	100	105	110	
Gly Lys Thr Gln Phe Cys Ile Met Met Ser Ile Leu Ala Thr Leu Pro	115	120	125	
Thr Asn Met Gly Gly Leu Glu Gly Ala Val Val Tyr Ile Asp Thr Glu	130	135	140	
Ser Ala Phe Ser Ala Glu Arg Leu Val Glu Ile Ala Glu Ser Arg Phe	145	150	155	160
Pro Arg Tyr Phe Asn Thr Glu Glu Lys Leu Leu Thr Ser Ser Lys	165	170	175	

FIG.1A

Val His Leu	Tyr Arg Glu Leu Thr	Cys Asp Glu Val Leu Gln Arg Ile	180	185	190
Glu Ser Leu	Glu Glu Ile Ile Ser Lys Gly Ile Lys Leu Val Ile				
195	200	205			
Leu Asp Ser Val Ala Ser Val Val Arg Lys Glu Phe Asp Ala Gln Leu					
210	215	220			
Gln Gly Asn Leu Lys Glu Arg Asn Lys Phe Leu Ala Arg Glu Ala Ser					
225	230	235			
Ser Leu Lys Tyr Leu Ala Glu Glu Phe Ser Ile Pro Val Ile Leu Thr					
	245	250			
Asn Gln Ile Thr Thr His Leu Ser Gly Ala Leu Ala Ser Gln Ala Asp					
260	265	270			
Leu Val Ser Pro Ala Asp Asp Leu Ser Leu Ser Glu Thr Ser Gly					
275	280	285			
Ser Ser Cys Val Ile Ala Ala Leu Gly Asn Thr Trp Ser His Ser Val					
290	295	300			
Asn Thr Arg Leu Ile Leu Gln Tyr Leu Asp Ser Ser Glu Arg Arg Gln Ile					
305	310	315			
Leu Ile Ala Lys Ser Pro Leu Ala Pro Phe Thr Ser Phe Val Tyr Thr					
	325	330			
Ile Lys Glu Gly Leu Val Leu Gln Ala Tyr Gly Asn Ser					
340	345	350			

FIG.1B

CGGACGCGTG GCGCGGGGA AACTGTGTAA AGGGTGGGA AACTTGAAG TTGGATGCTG
 CAGACCCGGC ATGGGTAGCA AGAAACTAAA ACGAGTGGGT TTATCACAAAG AGCTGTGTGA
 CCGTCTGAGT AGACATCAGA TCCTTACCTG TCAGGACTTT TTATGTCTTT CCCCACTGGA
 GCTTATGAAG GTGACTGGTC TGAGTTATCG AGGTGTCCAT GAACTTCTAT GTATGGTCAG
 CAGGCGCTGT GCCCCAAAGA TGCAAAACGGC TTATGGGATA AAAGCACAAA GGTCCTGCTGA
 TTCTTCACCA GCATTCTTAT CTACTACCTT TTCTGCTTTG GACGAAGCCC TGCATGGTGG
 TGTGGCTTGT GGATCCCTCA CAGAGATTAC AGGTCCACCA GGTGTGGAA AAATCAGTT
 TTGTATAATG ATGAGCATTT TGGCTACATT ACCCACCAAC ATGGGAGGAT TAGAAGGAGC
 TGTGGTGTAC ATTGACACAG AGTCTGCATT TAGTGCTGAA AGACTGGTTG AAATAGCAGA
 ATCCCCGTTT CCCAGATATT TTAACACTGA AGTCTACAA CTTTGTGACAA GTAGTAAAGT
 TCATCTTTAT CGGGAACCTCA CCTGTGATGA AGTTCTACAA AGGATTGAAT CTTTGGAAAG
 AGAAATTATC TCAAAAGGAA TTAACCTTGT GATTCTTGAC TCTGTTGCTT CTGTGGTCAG
 AAAGGAGTTT GATGCACAAC TTCRAGGCAA TCTCAAAGAA AGAAACAAGT TCTTGGCAAG
 AGAGGCATCC TCCTTGAAGT ATTTGGCTGA GGAGTTTCA ATCCCAGTTA TCTTGACGAA
 TCAGATTACA ACCCATCTGA GTGGAGCCCCT GGCTTCTCAG GCAGACCTGG TGTCTCCAGC

FIG.1C

TGATGATTG	TCCCTGTCTG	AAGGCACCTC	TGGATCCAGC	TGTGTGATAG	CCGCACTAGG	960
AAATACCTGG	AGTCACAGTG	TGAATACCCG	GCTGATCCTC	CAGTACCTTG	ATTCAGAGAG	1020
AAGACAGATT	CTTATTGCCA	AGTCCCTCT	GGCTCCCTTC	ACCTCATTTG	TCTACACCAT	1080
CAAGGAGGAA	GGCCTGGTTC	TTCAAGCCTA	TGGAAATTCC	TAGAGACAGA	TAAATGTGCA	1140
AACCTGTTCA	TCTTGCCAAG	AAAAATCCGC	TTTTCTGCCA	CAGAAACAAA	ATATTGGGAA	1200
AGAGTCTTGT	GGTGAAACAC	CCATCGTTCT	CTGCTAAAC	ATTTGGTTGC	TACTGTGTAG	1260
ACTCAGCTTA	AGTCATGGAA	TTCTAGAGGA	TGTATCTCAC	AAGTAGGATC	AAGAACAAGC	1320
CCAACAGTAA	TCTGCATCAT	AAGCTGATTT	GATACCATGG	CAC TGACAAT	GGGCACTGAT	1380
TTGATAACCAT	GGCACTGACA	ATGGGCACAC	AGGGAACAGG	AAATGGGAAT	GAGAGCAAGG	1440
GTTGGGTTGT	GTTCTGTGGA	CACATAGGTT	TTTTTTTTTA	ACTTTCTCTT	TCTAAAAATAT	1500
TTCATTTTGA	TGGAGGTGAA	ATTTATATAA	GATGAAATTA	ACCATTTTAA	AGTAAACAAT	1560
TCCGTGGCAA	CTAGATATCA	TGATGTGCAA	CCAGCATCTC	TGTCTAGTTC	CCAAATATTT	1620
CATCACCCCC	AAAAGCAAGA	CCCATAAACA	TTATGCAAGT	GTTCCCTATTT	CCCCCTCCTC	1680
CCAGCTCCTG	GGAAACCAAC	AATCTACTTT	TTTTCTATGG	CTTTACCTAA	TCTGGAAATT	1740
TCAAAATAAT	GGGATCAAAT	AGTTTCCCCA	AAAAAATAAA	AAAAAATAAA	AAAAAA	1797

FIG.1D

Met	Ser	Ser	Lys	Lys	Leu	Arg	Arg	Val	Gly	Leu	Ser	Pro	Glu	Leu	Cys
1			5						10					15	
Asp	Arg	Leu	Ser	Arg	Tyr	Leu	Ile	Val	Asn	Cys	Gln	His	Phe	Leu	Ser
		20						25					30		
Leu	Ser	Pro	Leu	Glu	Leu	Met	Lys	Val	Thr	Gly	Leu	Ser	Tyr	Arg	Gly
		35					40					45			
Val	His	Glu	Leu	Leu	His	Thr	Val	Ser	Lys	Ala	Cys	Ala	Pro	Gln	Met
		50				55					60				
Gln	Thr	Ala	Tyr	Glu	Leu	Lys	Thr	Arg	Arg	Ser	Ala	His	Leu	Ser	Pro
65					70					75				80	
Ala	Phe	Leu	Ser	Thr	Thr	Leu	Cys	Ala	Leu	Asp	Glu	Ala	Leu	His	Gly
				85					90					95	
Gly	Val	Pro	Cys	Gly	Ser	Leu	Thr	Glu	Ile	Thr	Gly	Pro	Pro	Gly	Cys
			100					105					110		
Gly	Lys	Thr	Gln	Phe	Cys	Ile	Met	Met	Ser	Val	Leu	Ala	Thr	Leu	Pro
		115					120					125			
Thr	Ser	Leu	Gly	Gly	Leu	Glu	Gly	Ala	Val	Val	Tyr	Ile	Asp	Thr	Glu
		130				135					140				
Ser	Ala	Phe	Thr	Ala	Glu	Arg	Leu	Val	Glu	Ile	Ala	Glu	Ser	Arg	Phe
145					150					155				160	
Pro	Gln	Tyr	Phe	Asn	Thr	Glu	Glu	Lys	Leu	Leu	Thr	Ser	Ser	Ser	Arg
				165					170					175	

FIG.1E

Val	His	Leu	Cys	Arg	Glu	Leu	Thr	Cys	Glu	Gly	Leu	Leu	Gln	Arg	Leu
			180					185					190		
Glu	Ser	Leu	Glu	Glu	Glu	Ile	Ile	Ser	Lys	Gly	Val	Lys	Leu	Val	Ile
		195					200					205			
Val	Asp	Ser	Ile	Ala	Ser	Val	Val	Arg	Lys	Glu	Phe	Asp	Pro	Lys	Leu
	210					215					220				
Gln	Gly	Asn	Ile	Lys	Glu	Arg	Asn	Lys	Phe	Leu	Gly	Lys	Gly	Ala	Ser
225					230					235					240
Leu	Leu	Lys	Tyr	Leu	Ala	Gly	Glu	Phe	Ser	Ile	Pro	Val	Ile	Leu	Thr
				245					250					255	
Asn	Gln	Ile	Thr	Thr	His	Leu	Ser	Gly	Ala	Leu	Pro	Ser	Gln	Ala	Asp
			260					265					270		
Leu	Val	Ser	Pro	Ala	Asp	Asp	Leu	Ser	Leu	Ser	Glu	Gly	Thr	Ser	Gly
		275					280					285			
Ser	Ser	Cys	Leu	Val	Ala	Ala	Leu	Gly	Asn	Thr	Trp	Gly	His	Cys	Val
	290					295					300				
Asn	Thr	Arg	Leu	Ile	Leu	Gln	Tyr	Leu	Asp	Ser	Glu	Arg	Arg	Gln	Ile
305					310					315					320
Leu	Ile	Ala	Lys	Ser	Pro	Leu	Ala	Ala	Phe	Thr	Ser	Phe	Val	Tyr	Thr
				325					330						
Ile	Lys	Gly	Glu	Gly	Leu	Val	Leu	Gln	Gly	His	Glu	Arg	Pro		
			340					345							
															350

FIG.1F

GGGAGCCCTG	GAAACATGAG	CAGCAAGAAA	CTAAGACGAG	TGGGTTTATC	TCCAGAGCTG	60
TGTGACCCGT	TAAGCAGATA	CCTGATTGTT	AACTGTCAGC	ACTTTTAAAG	TCTCTCCCCA	120
CTAGAACCTA	TGAAAGTGAC	TGGCCTGAGT	TACAGAGGTG	TCCACGAGCT	TCTTCATACA	180
GTAAGCAAGG	CCTGTGCCCC	GCAGATGCAA	ACGGCTATG	AGTTAAAGAC	ACGAAGGTCT	240
GCACATCTCT	CACCGGCATT	CCTGTCTACT	ACCCTGTGCG	CCTTGGATGA	AGCATTTGCAC	300
GGTGGTGTGC	CTTGTGGATC	TCTCACAGAG	ATTACAGGTC	CACCAGGTTG	CGGAAAAACT	360
CAGTTTGTGA	TAATGATGAG	TGTCCTTAGCT	ACATTACCTA	CCAGCCTGGG	AGGATTAGAA	420
GGGGCTGTGG	TCTACATCGA	CACAGAGTCT	GCATTTACTG	CTGAGAGACT	GTTTGAGATT	480
GCGGAATCTC	GTTTTCCACA	ATATTTTAAC	ACTGAGGAAA	AATTGCTTCT	GACCAGCAGT	540
AGAGTTTCATC	TTTGCCCGAGA	GCTCACCTGT	GAGGGGCTTC	TACAAAAGGCT	TGAGTCTTTG	600
GAGGAAGAGA	TCATTTTCGAA	AGGAGTTAAG	CTTGATGATTG	TTGACTCCAT	TGCTTCTGTG	660
GTCAGAAAAGG	AGTTTGACCC	GAAGCTTCAA	GGCAACATCA	AAGAAAGGAA	CAAGTTCTTG	720
GGCAAAAGGAG	CGTCCTTACT	GAAGTACCTG	GCAGGGGAGT	TTTCAATCCC	AGTTATCTTG	780
ACGAATCAAA	TTACGACCCA	TCTGAGTGGA	GCCCTCCCTT	CTCAAGCAGA	CCTGGTGTCT	840
CCAGCTGATG	ATTTGTCCCT	GTCGGAAGGC	ACTTCTGGAT	CCAGCTGTTT	GGTAGCTGCA	900
CTAGGAAACA	CATGGGGTCA	CTGTGTGAAC	ACCCGGCTGA	TTCTCCAGTA	CCTTGATTCA	960
GAGAGAAGGC	AGATTCTCAT	TGCCAAGTCT	CCTCTGGCTG	CCTTCACCTC	CTTTGTCTAC	1020
ACCATCAAGG	GGGAAGGCCT	GGTCTTCAA	GGCCACGAAA	GACCATAGGG	ATACTGTGAC	1080
CTTTGTCTAG	TGCTGATTGC	ATGTGACTCA	TGAAATGAAA	CAGGACTGCG	CTGCTTGGAA	1140
AAAGGAAACG	GAAGCCAACA	TAATGAGGAT	TAATTGGTTG	GTTGCTGTTG	AGGTGGTAAC	1200
AGTGATTTC	GACCCGGAAG	GTGAAGATGA	AGAAGCCTTT	ATCCAGTCTC	TGGATGCAGA	1260
GGCTAGGGGC	TCCACCACCG	TGGGATGTCA	GCGGCCATCG	TAATAAATTG	CACCTACACA	1320
AGCACCTTTC	AGCCATGCCC	CTCAAAGTGG	TTCAGCCACA	TTAATTAAAT	AAAGCCCACA	1380
ATCCCCCTAG	GGAGAGCAGG	AGGGGGACTA	ACAAGATTG	TAATTACAGA	AGGGAATAAT	1440
TCCGAATAAA	GTATTGTTCC	GCCAAAAAAA	AAAAAATAAA	AAAAAATAAA	AAAAAATAAA	1500
AAAAAATAAA	AAAAAATAAA	AAAAAATAAA	AAAAAATAAA	AAAAAATAAA	AAAAAATAAA	1525

FIG. 1G

MGSKKLKRUGLSQELCDRLSRHQILTCQDFLCSPLELMKDTGLS
NLS
YRGUHELLCMDSRACAPKMQTAYGIKAQRSADFSPAFLSTILSA
50
LDEALHGGVACGSLTEITGPPGCGKTQFCIMMSILATLPTNMGGL
100 A BOX
EGAUUYIDTESAFSAERLVEIAESRPRYPFNTTEKLLTSSKUHLY
150 P
RELTCDEVLQRIELEEIIISKGIKLWILDSVASURKEFDAQLQG
DNA 200 B BOX
NLKERNKFLAREASSLYLAEFEFSIPUILTNQITTHLSGALASQAD
250
LUSPADDLSLSEGTSGSSCVIAALGNWWSHUNTALILQYLDSEAR
300
QILIAKSPLAPFTSFUYTIKEEGLULQAYGNS*
350

FIG.2A

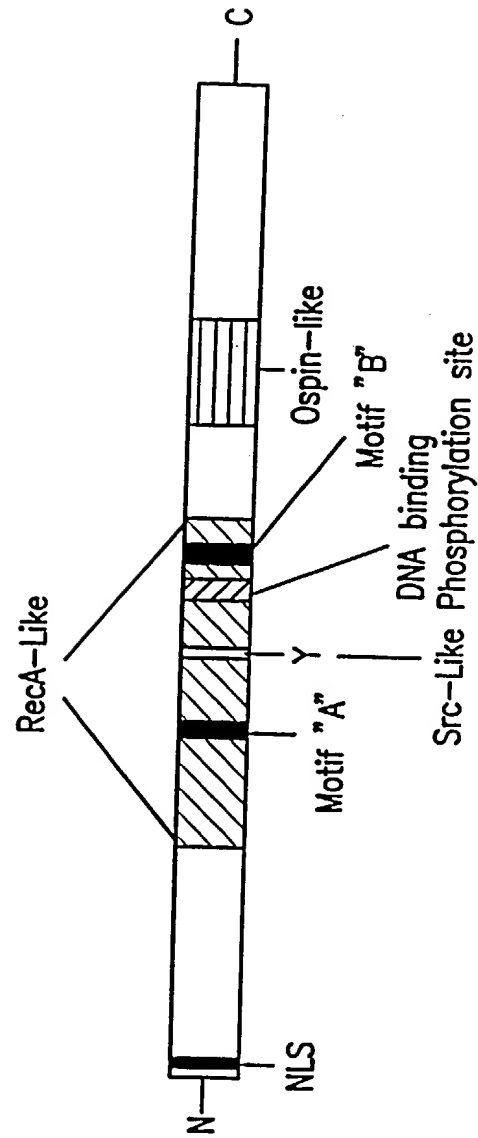


FIG.2B

U.m. 124 LNDARFASSCIVPPTQGYDGNFPGAQCFVYDS DAGSDSDARSSIDA VMHE 173
 Human 1 MGSKKLKR...VGLSQELCDRLSRHQILTCQDFLCLSPLELMKVTGLSYR 47
 174 DI.ELPSTFCRPQTPQTHDVARDEHHDGYLCDPKVDHASVARDVLSLGRQ 222
 48 GVHELLECMVSRA.....CAPKMQTAYGIKAQRSADFS 79
 223 RHVFSSGSRELDLLGGGVRSAVLTTELVGESGSGKTQMAIQVCTYAAIGL 272
 80 PAFLSTTSLDEALHGGVACGSLTETTGPFGCGKTQFCIMMSILATL.. 127
 273 VPLSQADDHDKGNNTFQSRTFVRDPIHASTKDDTLSDILQSYGMEPSIGS 322
 128PTNMGGLEG..... 136
 323 HRGMGACYITSGGERAAHSIVNRALELASFAINERFDRVYPVCDPTQSSQ 372
 137AVVYIDTESAFSAERLVEIA.....ESRFP RYF..... 164
 373 DADGRRDALLAKAQQLRGRQALANLHIACVADVEALEHALKYSLPGLIRR 422
 165 ...NTEEKLLLTSSKVHLYRELTCDEV..LQRIESLEEEI..... 199

FIG.2C

FIG. 2D

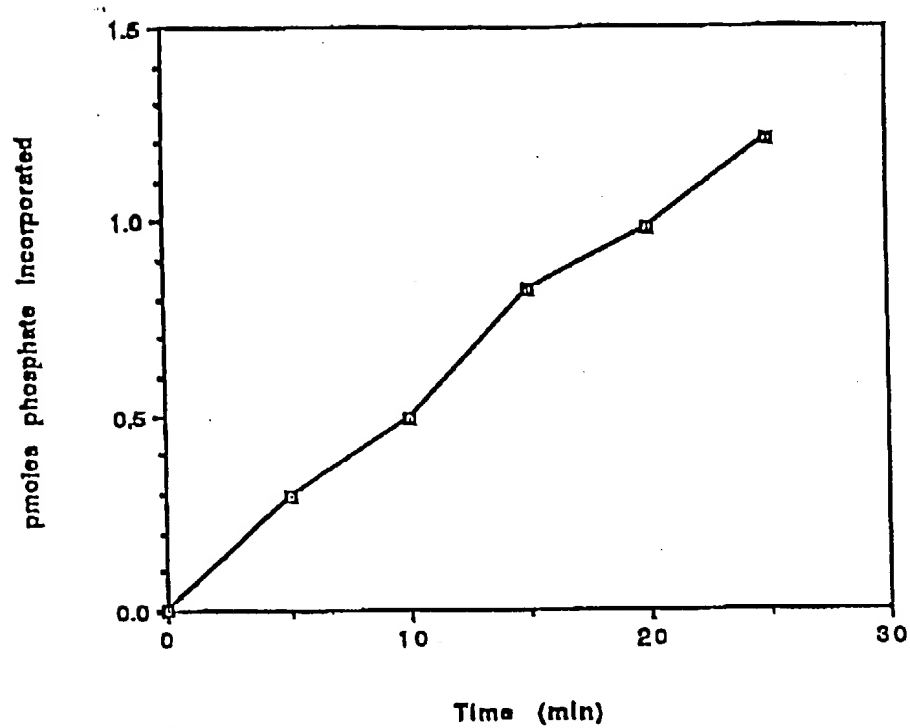


FIG. 3A

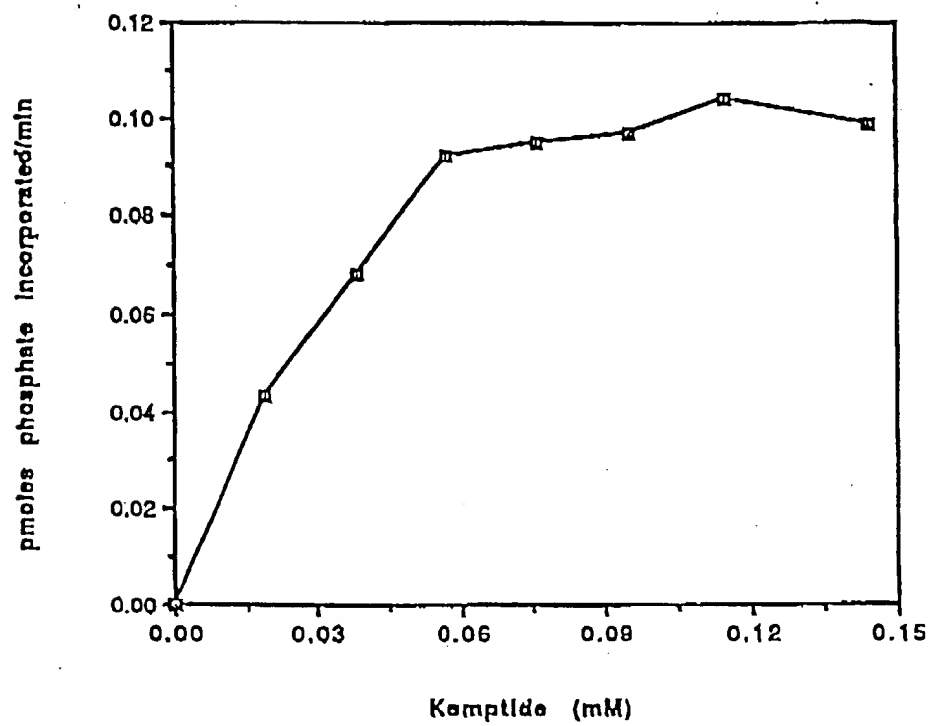


FIG. 3B

APPLICATION UNDER UNITED STATES PATENT LAWS

Atty. Dkt. No. PMS 263768
(M#)

Invention: Use of Chimeric Mutational Vectors to Change Endogenous Sequences in Solid Tissues

Inventors: Richard J. BARTLETT (Columbus, OH) and Thomas A. RANDO (Palo Alto, CA)

Pillsbury Madison & Sutro LLP
Intellectual Property Group
1100 New York Avenue, NW
Ninth Floor
Washington, DC 20005-3918
Attorneys
Telephone: (202) 861-3000

This is a:

- ☐ Provisional Application
- ☒ Regular Utility Application
- ☐ Continuing Application
- ☐ PCT National Phase Application
- ☐ Design Application
- ☐ Reissue Application
- ☐ Plant Application

- ☐ Substitute Specification
Sub. Spec. Filed _____
in App. No. _____ / _____
- ☐ Marked up Specification re
Sub. Spec. filed _____
In App. No. _____ / _____

SPECIFICATION

EXPRESS MAIL MAILING LABEL

Number: EL321554992US

Date of Deposit: May 20, 2000

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Commissioner of Patent and Trademarks, Washington DC 20231.

Gary Tanigawa, Reg. No. 43,180

ATTORNEY DOCKET NUMBER: 7991-086-99

SERIAL NUMBER: 09/685,403

REFERENCE: **AZ**

APPLICATION UNDER UNITED STATES PATENT LAWS

Atty. Dkt. No. PMS 263768
(M#)

Invention: Use of Chimeric Mutational Vectors to Change Endogenous Sequences in Solid Tissues

Inventors: Richard J. BARTLETT (Columbus, OH) and Thomas A. RANDO (Palo Alto, CA)

Pillsbury Madison & Sutro LLP
Intellectual Property Group
1100 New York Avenue, NW
Ninth Floor
Washington, DC 20005-3918
Attorneys
Telephone: (202) 861-3000

This is a:

- ☐ Provisional Application
- ☒ Regular Utility Application
- ☐ Continuing Application
- ☐ PCT National Phase Application
- ☐ Design Application
- ☐ Reissue Application
- ☐ Plant Application
- ☐ Substitute Specification
Sub. Spec Filed _____
in App. No. _____ / _____
- ☐ Marked up Specification re
Sub. Spec. filed _____
In App. No. _____ / _____

SPECIFICATION

EXPRESS MAIL MAILING LABEL

Number: EL321554992US

Date of Deposit: May 20, 2000

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Commissioner of Patent and Trademarks, Washington DC 20231.

Gary Tanigawa, Reg. No. 43,180

USE OF CHIMERIC MUTATIONAL VECTORS TO CHANGE ENDOGENOUS NUCLEOTIDE SEQUENCES IN SOLID TISSUES

CROSS-REFERENCES TO RELATED APPLICATIONS

5 This application claims priority benefit to provisional U.S. Appln. No. 60/135,139, filed May 21, 1999, and provisional U.S. Appln. No. 60/174,388, filed January 5, 2000, which are both incorporated by reference herein.

FIELD OF THE INVENTION

10 The invention concerns methods of treating genetic diseases or other pathologic conditions by making one or more specific changes in endogenous nucleotide sequences of solid tissues. These specific changes are mediated by oligonucleobases called chimeric mutational vectors (CMV). The CMV can be administered directly to the subject *in vivo*; in particular, the CMV can be injected into a solid tissue in which expression of the mutated gene
15 occurs. Such gene repair can reverse the disease or other pathologic condition caused by the mutation or, alternatively, can introduce a second change that compensates for the disease or condition causing mutation.

BACKGROUND OF THE INVENTION

20 The inclusion of a reference in this section is not to be understood as an admission that its teachings were publicly available prior to our invention of the subject matter disclosed herein or that they resulted from someone other than the inventors.

Chimeric Mutational Vector (CMV)

25 An oligonucleobase, which has complementary segments of deoxyribonucleotides and ribonucleotides, and contained a sequence homologous to a fragment of the bacteriophage M13mp19, has been described (Kmiec et al., Molecular and Cellular Biology 14:7163-7172, 1994). The oligonucleobase had a single contiguous segment of ribonucleotides. It is a substrate for the REC2 homologous pairing enzyme from *Ustilago maydis*. Thus, this enzyme
30 and the DNA mismatch repair machinery were suggested to be involved in gene repair.

Patent publication WO 95/15972, published June 15, 1995, and corresponding U.S. Appln. No. 08/353,657, filed December 9, 1994, now U.S. Patent No. 5,565,350 (the '350 patent) described chimeraplasts to genetically change eukaryotic cells. Examples with a

Ustilago maydis gene and the murine ras gene were reported. The latter example was designed to introduce a transforming mutation into the ras gene so that the successful mutation of the ras gene in murine NIH 3T3 cells would cause the growth of a colony of cells. The maximum rate of such transformation of cultured cells was less than 0.1%, i.e., less than 100 transformants per 10⁶ cells exposed to the CMV had a phenotype indicative of ras mutation. In the *Ustilago maydis* system, the rate of introduction of the genetic change was about 600 per 10⁶ cells. A chimeraplast was also designed to introduce a mutation into the human bcl-2 gene (Kmiec, Seminars in Oncology 23:188-193, 1996).

A chimeraplast was also designed to repair a mutation in codon 12 of K-ras (Kmiec, Advanced Drug Delivery Reviews 17:333-340, 1995). The chimeraplast was introduced into Capan 2, a cell line derived from a human pancreatic adenocarcinoma, using LIPOFECTIN cationic lipid. Twenty-four hours after the chimeraplasts were introduced, cells were harvested and genomic DNA was extracted. A fragment containing codon 12 of K-ras was amplified by PCR and the rate of conversion estimated by hybridization with allele-specific probes. The rate of repair was reported to be approximately 18%.

A chimeraplast has been designed to repair a mutation in the gene encoding liver/bone/kidney type alkaline phosphatase (Yoon et al., Proceedings of the National Academy of Sciences USA 93:2071-2076, 1996). The alkaline phosphatase gene was transiently introduced into CHO cells by a plasmid. Six hours later the chimeraplasts were introduced. The plasmid was recovered at 24 hours after introduction of the chimeraplast and analyzed. The results showed that approximately 30 to 38% of the alkaline phosphatase genes were repaired by the chimeraplast.

U.S. Appln. No. 08/640,517, filed May 1, 1996 and published as WO 97/41141, and Cole-Strauss et al., Science 273:1386-1389, 1996, disclose chimeraplasts that are used in the treatment of genetic diseases of hematopoietic cells, e.g., sickle cell disease, thalassemia, and Gaucher disease. U.S. Appln. No. 08/664,487, filed June 17, 1996 and published as WO 97/04871, describes chimeraplast having non-natural nucleotides for use in specific, site-directed mutagenesis. The chimeraplasts described in the applications and publications of Kmiec and his colleagues contain a central segment of DNA:DNA homoduplex and flanking segments of RNA:DNA heteroduplex or 2'-O-Me-RNA:DNA heteroduplex. Kren et al., Hepatology 25:1462-1468, 1997, report the successful use of a CMV in non-replicating primary hepatocytes.

U.S. Appln. No. 60/054,837, filed August 5, 1997, U.S. Appln. No. 09/108,006, filed June 30, 1998, and U.S. Appln. No. 60/064996, filed November 5, 1997, concern the use of chimeraplasts in non-replicating cells and compositions of CMV and macromolecular carriers, including macromolecular carriers that have ligands for clathrin-coated pit receptors.

5

Introduction of DNA into Muscle Cells

There are several references that report the introduction and expression of plasmid DNA encoding the dystrophin protein into skeletal muscle (Acsadi et al., Nature 352:815-818, 1991; Danko et al., Human Molecular Genetics 2:2055-2061, 1993; Bartlett et al., Cell Transplantation 5:411-419, 1996; Wells et al., FEBS Letters 332:179-182, 1993; Fritz et al., Pediatric Research. 37:693-700, 1995; Wolff et al., Human Molecular Genetics 1:363-369, 1992; Inui et al., Brain & Development 18:357-361, 1996). A general method of introducing DNA into a muscle cell for the purpose of inducing an immune response in a host is disclosed in U.S. Patent Nos. 5,589,466 and 5,580,859. The expression of an exogenous dystrophin gene is an example in these patents.

Experiments directed at determining a ligand that can be used to introduce large DNA fragments into the myofibers of DMD patients have been reported (Feero et al., Gene Therapy 4:664-674, 1997). The use of liposomes to deliver DNA to myofibers for expression without the use of a targeting ligand has also been described (Templeton et al., Nature Biotechnology 15:647-652, 1997).

20

Molecular Biology of Muscular Dystrophy

The muscular dystrophies comprise a genetically and clinically diverse set of diseases characterized by abnormalities of the skeletal muscle (reviewed by Straub et al., Current Opinion in Neurology 10:168-175, 1997). The muscular dystrophies can be classified by the mode of inheritance, i.e., autosomal dominant, autosomal recessive, and X-linked, and each type further divided according to the chromosomal locus and even the effected gene, if known.

25

The most common muscular dystrophy is X-linked with the dystrophin gene effected. The dystrophin gene occupies 2,300 kb or about 1.5 % of the X-chromosome. Its mature transcript is 14 kb and encodes a protein of 3685 amino acids having a molecular weight of 427 kd. The gene contains 79 exons. The dystrophin gene is extraordinarily large; it is about half the size of an *E. coli* genome. There is no clear explanation for its size. See Worton &

30

The dystrophin protein contains an N-terminal binding region, that binds to intracellular filamentous actin (which is not the actin of the contractile apparatus), a C-terminal binding domain that binds to a transmembranous glycoprotein complex which in turn binds to laminin, and a connective region. Under physiologic conditions, dystrophin exists as a homodimer and connects the actin filaments with the glycoprotein complex as well as linking each.

Although there are multiple mutations of dystrophin that result in muscular dystrophy, the mutations can be classified into types. The milder form, termed Becker Muscular Dystrophy (BMD), is associated with genomic deletions or mRNA processing errors that do not alter the reading frame of the mature mRNA and, hence result in a mutant protein that contains intact N-terminal and C-terminal binding domains. In the more severe form, termed Duchenne Muscular Dystrophy (DMD), the dystrophin protein lacks a C-terminal binding domain and is usually unstable. DMD typically results from point mutations that introduce in-frame termination codons or from insertion or deletion mutations that result in a frame-shift. See, generally, Koenig et al., *American Journal of Human Genetics* 45:498-506, 1989; Prior et al., *Human Mutation* 5:263-268, 1995; Koenig et al., *Cell* 50:509-517, 1987; Baumbach et al., *Neurology* 39:465-474, 1989.

The relationship between the pathophysiology of DMD and BMD and the physiologic function of dystrophin is complex. Dystrophin is not required to transmit the force of the contractile apparatus to the tendonous connections of the muscle. Rather, the defective muscles undergo an ongoing series of focal necrosis of the myofibers, which ultimately exceed the repair capacity of the muscle. The end stage disease is characterized by fibrosis between myofibers, atrophy, and weakness.

Dystrophin Replacement Gene Therapy

Several groups have attempted to treat DMD by introducing genes encoding dystrophin into the myofibers of affected individuals. A variety of methods have been employed and can be classified into three groups: *in situ* replacement gene therapy; *ex vivo* replacement gene therapy using autologous myoblasts, which are then reimplanted; and allogenic transplantation of wild-type myoblasts.

Examples of the first type include the aforementioned transfections of differentiated myofibers using DNA and non-biologic carriers. This form of therapy has been of limited

value because of the low efficiency of transfection. The use of adenovirus based vectors to increase efficiency has been reported. See, generally, Vincent et al., *Nature Genetics* 5:130-134, 1993; Ragot et al., *Gene Therapy* 1 Suppl 1:S53-S54, 1994; Acsadi et al., *Human Gene Therapy* 7:129-140, 1996; Deconinck et al., *Proceedings of the National Academy of Sciences USA* 93:3570-3574, 1996; Clemens et al., *Gene Therapy* 3:965-972, 1996; Haecker et al., *Human Gene Therapy* 7:1907-1914, 1996; Chen et al., *Proceedings of the National Academy of Sciences USA* 94:1645-1650, 1997; Yang et al., *Journal of Virology* 69:2004-2015, 1995; Haecker et al., *Human Gene Therapy* 7:1907-1914, 1996. Although efficiencies as high as 50% have been reported in experimental animal systems (Ragot et al., *Nature* 361:647-650, 1993), adenovirus-based therapies have likewise been of limited value to date because the expression of dystrophin has been transient and there is an immune response to the adenovirus vector that limits the possibilities of repeated therapy. Although such gene therapy has not proved to be a practical clinical modality, it has been useful to demonstrate that the expression of a wild-type dystrophin in an DMD model system results in amelioration of the disease (Danko et al., *Human Molecular Genetics* 2, 2055-2061, 1993).

Techniques for the culture of myoblasts from normal individuals have been reported (U.S. Patent No. 5,538,722). Dystrophin has been transferred into cultured myoblasts (Dunckley et al., *FEBS Letters* 296:128-134, 1992) but this approach has not been pursued because a secondary effect of DMD is a decline in the numbers of myoblasts that can be recovered in culture (Webster & Blau, *Somatic Cell & Molecular Genetics* 16:557-565, 1990).

Successful engraftment of allogenic cultured myoblasts has been reported (U.S. Patent No. 5,130,141; Law et al., *Cellular Transplantation* 1:235-244, 1992). Other studies, however, have failed to confirm the clinical benefit of allogenic myoblast grafts under controlled conditions (Gussoni et al., *Nature* 356:435-438, 1992; Karpati et al., *Annals of Neurology* 34:8-17, 1992). There is consequently a need for a therapy that results in the long-term expression of functional dystrophin in muscle fibers affected by muscular dystrophy. Ideally, the therapy should be applicable to all solid tissues whether or not they are highly vascularized.

A further limitation of both myoblast engraftment and non-viral gene therapy is a requirement for local delivery, such that multiple injections are required to treat even a single large muscle and obtain permanent effects (e.g., gene repair). Reports to the contrary with regard to myoblast engraftment notwithstanding (e.g., Hughes & Blau, *Nature* 345:350-353, 1990; Neumeyer et al., *Neurology* 42:2258-2262, 1992), more recent studies have not confirmed that transvascular engraftment into muscle fibers occurs to any practical extent.

Two well-characterized animal models exist for Duchenne muscular dystrophy, the *mdx* mouse (Bulfield et al., Proceedings of the National Academy of Sciences USA 81:1189-1192, 1984; Sicinski et al., Science 244:1578-1579, 1989) and the golden retriever dog (Kornegay et al., Muscle and Nerve 11:1056-1064, 1988; Sharp et al., Genomics 13:115-121, 1992). In both cases, a point mutation has been identified as causing disease: the mouse having a nonsense mutation in exon 23 and the dog having a splice acceptor site mutation in intron 6 causing a frame-shift due to complete deletion of exon 7 from the mature canine dystrophin mRNA (Wilton et al., Muscle and Nerve 20:728-734, 1997; Wilton et al., Neuromuscular Disorders 7:329-335, 1997; Schatzberg et al., Muscle and Nerve 21:991-998, 1998). Alternate splicing mechanisms, which restore the dystrophin reading via removal of mutation containing out-of-frame exons, have been suggested to play a causal role for the presence of dystrophin positive staining "revertant fibers" in both models, although no evidence of true reversion of these point mutations at the genomic level have been reported. A considerable amount of effort has gone into the study of gene therapy in the *mdx* model using direct DNA injection (Acsadi et al., Nature 352:815-818, 1991) viral vectors (Danko et al., Human Molecular Genetics 2:2055-2061, 1993; Wells et al., FEBS Letters 332:179-182, 1992) and myoblast transplantation (Fritz et al., Pediatric Research 37:693-700, 1995; Inui et al., Brain & Development 18:357-361, 1996) with modest levels of short-term success due to limitations of transfection targeting and efficiency, and either acute or chronic immune responses directed against cells which express the therapeutic gene product (Kinoshita et al., Acta Neuropathologica 91:489-493, 1996; Kinoshita et al., Neuromuscular Disorders 6:187-193, 1996; Yang et al., Journal of Virology 70:7209-7212, 1996; Yang et al., Gene Therapy 3:137-144, 1996; Worgall et al., Human Gene Therapy 8:37-44, 1997). Recent studies have suggested that myoblast transplantation therapy of Duchenne muscular dystrophy is also ineffective (Partridge et al., Nature Medicine 4:1208-1209, 1998; Mendell et al., New England Journal of Medicine 333:832-838, 1995). Long-term correction of dystrophin deficiency requires a permanent effect such as gene repair which will provide stable expression of dystrophin without the problems associated with therapies such as delivery of expression vectors, viruses, and cell implantation.

Recently, a novel chimeric RNA and DNA oligonucleotide (i.e., a type of chimeraplast) was used to correct the sickle-cell globin allele in a lymphoblast cell line (Cole-Strauss et al., Science 273:1386-1389, 1996). This technique, termed chimeraplasty, is believed to rely on regions of sequence homology (i.e., mutator regions) designed into the chimeraplast that brackets the site of the chromosomal mutation and directs the host cell DNA mismatch repair

mechanism to correct the endogenous sequence to that designated within the mutator region (Ye et al., Molecular Medicine Today 4:431-437, 1998). In the sickle cell study, this resulted in the correction to the wild-type nucleotide sequence of 20% of the chromosomes bearing the sickle-cell globin mutation.

5 A critical issue in the field of gene therapy is reliable and safe introduction of nucleic acid into the subject's cells. Introduction of large, highly charged molecules (e.g., expression vectors used in gene therapy) has proved challenging, and current protocols have been very limited and generally laborious. Thus, we show that chimeric mutational vectors and direct injection into solid tissue affected by a genetic mutation improves the efficiency of gene repair
10 in well-characterized animal models of a human genetic disease. In particular, products and processes effective for introducing the chimeric mutational vector into cells of skeletal muscle (e.g., myoblasts, myocytes, myotubes, myofibers), and thereby correct dystrophin mutations therein, are provided. Similar products and processes are envisioned for other inherited and acquired genetic mutations. Other advantages of the invention beside those noted above will
15 be appreciated by a person skilled in the art from the description below.

SUMMARY OF THE INVENTION

A composition is provided that includes at least one chimeric mutational vector (CMV). Methods of making and using such compositions, which are used to change an endogenous
20 nucleotide sequence of an affected cell in solid tissue and thereby correct a genetic mutation that causes a disease or other pathologic condition, are also provided.

Introducing at least one chimeric mutational vector (CMV) can mediate one or more sequence-specific changes in the endogenous sequence of at least some cells of the solid tissue. Applications of this invention are not limited to repair of a gene's coding sequences because
25 non-coding and other chromosomal sequences could also be changed. For example, point mutations (e.g., nonsense or missense changes) and frame-shift mutations (e.g., insertions or deletions) in the coding region of a gene could be repaired, as well as genetic mutations in transcriptional regulatory regions (e.g., promoter, silencer, enhancer), initiation and termination sites for transcription or translation, or splice donors/acceptors.

30 We illustrate the operation of the invention by correction of dystrophin mutations in skeletal muscle. But more generally, any disease or other pathologic condition could be treated if the genetic basis was known: e.g., factor VIII and factor IX of liver for hemophilia A and B, respectively; UDP-glucuronosyltransferase of liver for Crigler-Najjar syndrome; expression of

tyrosine hydroxylase or other enzymes involved in L-dopamine biosynthesis could be increased in the substantia nigra to treat Parkinson's disease. Other mutated genes in liver which could be changed by this invention are also known to cause familial hypercholesterolemia, mucopolysaccharidosis, familial amyloidosis, phenylketonuria, maple syrup urine disease, hemochromatosis, α 1-antitrypsin deficiency, Wilson's disease, and ornithine transcarbamylase deficiency. Moreover, beneficial mutations could be made in a "normal" gene to prevent disease: e.g., APOB 100 may could be truncated or APO A1 may be altered to the Milano allele to increase serum high-density lipoproteins (HDL), and thereby reduce the circulating amount of low-density lipoproteins (LDL). See Scriver et al. (eds.), *Metabolic Basis of Inherited Disease*, McGraw-Hill (New York, NY, 1993) and Online Mendelian Inheritance in Man, OMIM database, Center for Medical Genetics, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD) at <http://www.ncbi.nlm.nih.gov/Omim/> for further information on human diseases and pathologic conditions for which genes and mutations have been identified. Mutations in oncogenes and tumor suppressor genes could also be repaired to treat neoplastic disease (e.g., cell cycle regulatory genes, DNA repair gene). For example, it might be possible to treat cancers of the muscle (e.g., sarcoma), liver (i.e., hepatoma), skin (e.g., melanoma), or brain (e.g., glioblastoma).

Gene repair is a process by which a specific alteration is introduced into an existing gene of a cell of the subject suffering from a disease. Gene repair differs from gene therapy in that gene therapy introduces an exogenous DNA fragment into the genome of a cell that is then expressed as the protein encoded by the introduced fragment. Gene repair, however, directs the DNA repair process of the subject cell to introduce the desired, specific alteration into the genome of the host cell. CMV does not need to be transcribed into an RNA transcript and does not have to encode a functional protein. This invention is based on the discoveries that CMV can be efficiently introduced into cells of solid tissues and that their nuclei are able to effect gene repair. Thus, delivery of a CMV into a cell is able to mediate a specific sequence change at high efficiency *in vivo*, and without the need for *in vitro* tissue culture or selection.

The sequence-specific genetic alteration can be made using a CMV as in "naked" form or in a delivery vehicle. Transfection agents such as, for example, lipids, viral particle, salt and polymeric precipitants, etc., may or may not be used to aid the introduction of the CMV into at least some cells of the solid tissue. Furthermore, the CMV may or may not be complexed with a macromolecular carrier to which is attached a specific ligand, e.g., a glucosyl moiety. The

ligand may also be selected to bind to a cell-surface receptor that is internalized into the cell through clathrin-coated pits into endosomes. Alternatively, the CMV may be linked directly to the ligand without employing an intermediate macromolecular carrier. Targeted delivery of the CMV may also be achieved by using a ligand for a cellular receptor found specifically on the target tissue which is endocytosed. Other tissues which may be targeted include nervous tissues (e.g., brain, eye, central and peripheral nerves, glia); hematopoietic tissues (e.g., bone marrow, liver); reproductive tissues and glands (e.g., breast, adrenal gland, pituitary gland, thyroid gland); connective tissues, smooth muscle, striated muscle (e.g., skeletal, heart), and skin; and other solid tissues. Another optional additive is one that can be used to indicate the injection track of the composition in a treated solid tissue.

In alternative embodiments, the invention concerns the *ex vivo* use of gene repair to correct genetic mutations in cultured autologous cells of the solid tissue, which can then be engrafted into a subject. Furthermore, *in utero* use of gene repair may correct mutations prior to development of symptoms and when the number of cells in the solid tissue is reduced. Expansion of cells whose genetic mutations have been corrected because of a selective growth advantage conferred by the functional gene and/or by induction of regeneration (e.g., barium chloride for muscle) can be used to increase the proportion of cells in the solid tissue that have undergone gene repair.

Our invention is described below and its advantages over the prior art are illustrated by way of those particular embodiments and certain technical features.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic of a chimeric mutational vector (CMV).

Figure 2 shows the normal human nucleotide sequence (SEQ ID NO:1), the normal canine nucleotide sequence (SEQ ID NO:2), the GRMD mutant nucleotide sequence (SEQ ID NO:3), and the nucleobase sequence of the CMV used for repair of the GRMD mutation (SEQ ID NO:4). The CMV sequence has a two-base mismatch as compared to the canine sequence designed to help distinguish both mutant and wild-type sequences from the repaired sequence.

Figure 3 shows a timeline for injections (dark vertical arrow and horizontal line for left limb treatment and cross-hatched vertical arrow and horizontal line for right limb treatment) and biopsies. The elapsed time until necropsy was 48 weeks for the left limb and 39 weeks for the right limb.

Figure 4 shows the locations of primers and mutations in the canine dystrophin gene.

Figure 5 shows a normal nucleotide sequence (SEQ ID NO:5) and the *mdx* mutation (SEQ ID NO:6) in panel A, the design of chimeric mutational vector MDX1 designed to repair the *mdx* mutation in a murine dystrophin gene (SEQ ID NO:7) in panel B, and a putative mechanism for gene repair to produce a corrected *mdx* allele (SEQ ID NO:8) in panel C.

5

DETAILED DESCRIPTION OF THE INVENTION

Multiple lines of evidence confirm that direct *in vivo* injection into dystrophic skeletal muscle of an appropriately designed and synthesized chimeric oligonucleobase (i.e., a chimeric mutational vector or CMV) results in reversion of the genetic mutation causing GRMD in dogs and the *mdx* mutation in mice. It is envisioned that such CMV-mediated gene repair can also be accomplished in humans having Duchenne and Becker muscular dystrophy. We have also surprisingly found that use of a lipid carrier vehicle to introduce the CMV into cells with a dystrophin mutation was required in dogs for sustained expression of corrected dystrophin transcripts, while successful gene repair of a point mutation in mice was not so limited.

10 In accordance with these teachings, those skilled in the art will appreciate that the invention can be used to treat muscular dystrophies caused by mutations in genes other than dystrophin. For example, the invention can also be used to correct mutations in Emery-Dreifuss muscular dystrophy caused by mutations in emerin, an X-linked gene, and recessive limb-girdle muscular dystrophy caused by mutations in the sarcoglycan genes, which are encoded on autosomes.

15 Figure 1 shows a diagram of a CMV according to one embodiment of the invention. Segments "a" and "c-e" are target gene specific segments of the CMV. The sequences of segment "a" and "c-e" are complements of each other. The sequence of segments "f" and "h" are also complements of each other but are unrelated to the specific target gene and are selected merely to ensure the stability of hybridization in order to protect the 3' and 5' ends. Additional protection of the 3' and 5' ends can be accomplished by making the 5' and 3' most internucleobase bonds a phosphorothioate, phosphonate or any other nuclease-resistant bond. The sequence of segments "f" and "h" can be 5'-GCGCG-3' or permutations thereof. Segments "g" and "b" can be any linker that covalently connects the two strands, e.g., four unpaired nucleotides or an alkoxy oligomer such as polyethylene glycol. When segments "g" and "b" are composed of other than nucleobases, then segments "a", "c-f" and "g" are each an oligonucleobase chain. The ribo-type nucleobase segments are segments "c" and "e," which form hybrid-duplexes by Watson-Crick base pairing to the complementary portions of segment

“a.” The segment “a” can have the sequence of either the coding or non-coding strand of the gene.

The sequence of the CMV useful to treat a particular subject depends upon the location and type of the mutation of the subject. Mutations consisting of the replacement of a single
5 base that causes a premature in-frame termination codon, can be treated by CMV comprising the sequence of the wild-type gene at the locus of the mutation. As used herein, a CMV has a particular sequence if either strand of the CMV comprises the sequence or comprises a sequence containing ribo-type nucleobase equivalents with uracil bases replacing thymine bases. A frame-shifting deletion of a fragment of an exon or even of a complete exon can be
10 treated by a CMV that differs from the mutated sequence by the presence of a one or two base insertion or deletion such that the correct reading frame is restored downstream of the mutation. Depending on the size of the deletion, gene repair can restore some or all of the normal function of dystrophin in the affected cell. A single-base substitution that affects the splicing of the dystrophin message can be similarly repaired to result in functional dystrophin.

15 Techniques for the diagnosis of DMD and BMD, as well as the localization and identification of the mutation in the human dystrophin gene responsible for disease, are well known to those skilled in the art. These techniques include the use of antibodies specific to the amino and carboxyl terminals of dystrophin (Bulman et al., American Journal of Human Genetics 48:295-304, 1991; Arahata et al., Journal of Neurological Science 101:148-156,
20 1991). Such antibody preparations in combination with western blotting can be used to distinguish internal deletions and point mutation that effect reading frame from deletion mutations that do not. The use of RT-PCR with mixtures of multiple exon specific primers that produce PCR fragments of distinguishable diagnostic size allows for the rapid detection of exon deletions in a subject's dystrophin mRNA (Abbs et al., Journal of Medical Genetics
25 28:304-311, 1991; Beggs et al., Human Genetics 86:45-48, 1990). The sensitivity of RT-PCR diagnosis is sufficient to permit the analysis of dystrophin message from peripheral blood, and identification of the mutation by the sequencing of the product (Roberts, American Journal of Human Genetics 49:298-310, 1991).

The sequence of the homologous region of a CMV of the invention can be selected in
30 accordance with the mutation's location or by the location that is selected for an insertion or deletion to restore the reading frame of the gene. The sequence of the homologous region will have the sequence or its equivalent of a fragment of an exon or an intron that is located within about 25 nucleotides of the exon or of a fragment that bridges an intron and an exon. As used

herein the term "flanking intron" refers to the 21 nucleotides of the intron adjacent to an exon. The nucleotide sequence of the exons and flanking intron sequences of the human dystrophin gene are known. Intron sequences not yet published can be obtained by standard techniques well known to those skilled in the art, using the sequence of the exon and the knowledge of the restriction map of the dystrophin gene (the size of the genomic Hind III fragment containing each exon of the dystrophin gene is disclosed in Roberts et al., Genomics 16:536-538, 1993).

CMV may be introduced into solid tissues by intravenous or intraarterial routes for those that are extensively vascularized. Preferred transfection methods, however, involve direct administration to the affected solid tissue that do not deliver the CMV throughout the system in significant amounts. This localizes gene repair to places where it will result in effective treatment while reducing the amount of CMV that is expended and minimizing effects in cells unaffected by the genetic disease or other pathologic condition. Such techniques may include biolistics and electroporation, but direct injection by hypodermic needle is preferred. In particular, administration of a composition localized to affected parenchyma or interstitial spaces proximal to affected tissue are preferred. Alternative techniques include sustained infusion of affected solid tissue by permeable matrices or pumps. Direct administration to localized spaces can be monitored in real time by including an indicator in the composition or determining its distribution at later times.

Methods of treatment according to the invention administer CMV alone or with other agents in a composition in effective amounts. Such treatment of mammalian subjects in need thereof may be (a) therapeutic to treat existing disease and other pathologic conditions and/or (b) prophylactic to prevent or at least reduce the propensity of developing disease and other pathologic conditions. Therapeutically or prophylactically effective amount, as recognized by those of skill in the art, will be determined on a case by case basis. Factors to be considered include, but are not limited to: the tissue-type of the targeted cell and its ability to replicate, synapse, or recombine nucleic acids, the genetic sequence to be altered, the disease or other condition to be treated, and the medical history and status of the subject to be treated. For example, acquired mutations may result in sporadic disease and other pathologic conditions that are easier to treat because gene repair is required in only a few cells.

Chimeric Mutational Vectors (CMV)

Compositions containing at least one chimeric mutational vector (CMV) may be used to deliver the CMV into muscle cells, at least some of which will target the dystrophin gene and

direct sequence-specific alterations therein (e.g., insertions, deletions, substitutions of one to six bases). A duplex oligonucleobase consisting of more than 200 deoxyribonucleotides and no nucleotide derivatives is not considered a CMV. Typically, a CMV is characterized by being a duplex oligonucleobase, including ribo-type and deoxyribo-type nucleobases, of lengths between about 20 and about 120 nucleobases or equivalently between about 10 and about 60 Watson-Crick nucleobase pairs.

“Chimeric mutational vectors” are described in U.S. Patent No. 5,565,350 as a duplex mixed oligonucleobase, which contains at least one strand of ribo-type and deoxyribo-type nucleobases, hybridized to each other. At least one region of contiguous unpaired nucleobases is disposed so that the unpaired region separates the oligonucleobase into a first strand and a second strand. The region of contiguous unpaired nucleobases connects a region of Watson-Crick paired nucleobases of at least 15 base pairs in length, in which the first strand’s nucleobases are complementary to the second strand’s nucleobases. The first strand may comprise a region of at least three to nine contiguous nucleobases comprised of a 2’-O or 2’-O-Me ribose, which form a hybrid-duplex within the region of Watson-Crick paired bases. Two regions homologous with the sequence of the target gene flank a heterologous region with the alteration. The second strand may contain no nucleobases comprised of a 2’-O or 2’-O-Me ribose. In such a CMV, the first strand may comprise two regions of nucleobases comprised of a 2’-O- or 2’-O-Me ribose that form two regions of hybrid-duplex, each hybrid-duplex region having at least four or eight base pairs of length, and an interposed region of at least four or eight base pairs of homo-duplex disposed between the hybrid duplex regions. The interposed region of homo-duplex may consist of between four and 50, or between 30 and 1,000, 2’-deoxyribose base pairs.

“Oligonucleobases” are polymers of nucleobases, which polymer can hybridize by Watson-Crick base pairing to a DNA having the complementary sequence. Nucleobases comprise a base, which is a purine, pyrimidine, or a derivative or analog thereof. Nucleobases include peptide nucleobases, the subunits of peptide nucleic acids, and morpholine nucleobases as well as nucleobases that contain a pentosefuranosyl moiety (e.g., a substituted riboside or 2’-deoxyriboside). A “nucleobase” contains a base, which is either a purine or a pyrimidine or analog or derivative thereof. There are two types of nucleobases. Ribo-type nucleobases are ribonucleosides having a 2’-hydroxyl, substituted 2’-hydroxyl or 2’-halo-substituted ribose. All nucleobases other than ribo-type nucleobases are deoxyribo-type nucleobases. Thus, deoxy-type nucleobases include peptide nucleobases.

“Nucleosides” are nucleobases attached to a pentosefuranosyl sugar, e.g., an optionally substituted riboside or 2'-deoxyriboside. Nucleosides can be linked by one of several linkages, which may or may not contain a phosphorus, including substituted phosphodiester bonds (e.g., phosphorothioate or triesterified phosphates). Nucleosides that are linked by unsubstituted phosphodiester bonds are termed nucleotides. Other types of heteroatom linkages contain a nitrogen, sulfur, or oxygen.

A oligonucleobase compound has 5' and 3' end nucleobases, which are the ultimate nucleobases of the polymer. Nucleobases are either deoxyribo-type or ribo-type. Ribo-type nucleobases are pentosefuranosyl containing nucleobases wherein the 2' carbon is a methylene substituted with a hydroxyl, substituted oxygen or a halogen. Deoxyribo-type nucleobases are nucleobases other than ribo-type nucleobases and include all nucleobases that do not contain a pentosefuranosyl moiety (e.g., peptide nucleic acids).

An oligonucleobase strand generically includes regions or segments of oligonucleobase compounds that are hybridized to substantially all of the nucleobases of a complementary strand of equal length. An oligonucleobase strand has a 3' terminal nucleobase and a 5' terminal nucleobase. The 3' terminal nucleobase of a strand hybridizes to the 5' terminal nucleobase of the complementary strand. Two nucleobases of a strand are adjacent nucleobases if they are directly covalently linked or if they hybridize to nucleobases of the complementary strand that are directly covalently linked. An oligonucleobase strand may consist of linked nucleobases, wherein each nucleobase of the strand is covalently linked to the nucleobases adjacent to it. Alternatively a strand may be divided into two chains when two adjacent nucleobases are unlinked. The 5' (or 3') terminal nucleobase of a strand can be linked at its 5'-O (or 3'-O) to a linker, which linker is further linked to a 3' (or 5') terminus of a second oligonucleobase strand, which is complementary to the first strand, whereby the two strands form one oligonucleobase compound. The linker can be an oligonucleobase, an oligonucleobase or other compound. The 5'-O and the 3'-O of a 5' end and 3' end nucleobase of an oligonucleobase compound can be substituted with a blocking group that protects the oligonucleobase strand. Of course, closed circular oligonucleotides do not contain 3' or 5' end nucleotides. Note that when an oligonucleobase compound contains a divided strand, the 3' and 5' end nucleobases are not the terminal nucleobases of a strand.

As used herein the terms 3' and 5' have their usual meaning. The terms “3' most nucleobase,” “5' most nucleobase,” “first terminal nucleobase,” and “second terminal nucleobase” have special definitions. The 3' most and second terminal nucleobase are the 3'

terminal nucleobases, as defined above, of complementary strands of a recombinagenic oligonucleobase. Similarly, the 5' most and first terminal nucleobase are 5' terminal nucleobases of complementary strands of a recombinagenic oligonucleobase.

5 More generally, the CMV is a polymer of nucleobases, which polymer hybridizes (i.e., form Watson-Crick base pairs of purines and pyrimidines) in a duplex structure. Each CMV can be divided into a first and a second strand of at least 12 nucleobases and not more than 75 nucleobases. The length of the strands may be each between 20 and 50 nucleobases. The strands contain regions that are complementary to each other. The two strands may be complementary to each other at every nucleobase except the nucleobases wherein the target
10 sequence and the desired sequence differ. At least two non-overlapping regions of at least five nucleobases are preferred.

If the strands are complementary to each other at every nucleobase, the sequence of the first and second strands consists of at least two regions that are homologous to the target gene and one or more regions (the "mutator regions") that differ from the target gene and introduce
15 the genetic change into the target gene. The mutator region is directly adjacent to homologous regions in both the 3' and 5' directions. The two homologous regions may be at least three, six, or 12 nucleobases in length. The total length of all homologous regions may be at least 12, between 16 and 60, or between 20 and 60 nucleobases in length. The total length of the homology and mutator regions together may be between 25 and 45, between 30 and 45, or
20 between 35 and 40 nucleobases. Each homologous region can be between eight and 30, between eight and 15 nucleobases, or about 12 nucleobases long. The mutator region may consist of 20 or fewer, six or fewer, or three or fewer nucleobases. The mutator region can be of a length different than the length of the sequence that separates the regions of the target gene homology with the homologous regions of the CMV so that an insertion or deletion of the
25 target gene results. When the CMV is used to introduce a deletion in the target gene there is no nucleobase identifiable as within the mutator region. Rather, the mutation is effected by the juxtaposition of the two homologous regions that are separated in the target gene. The length of the mutator region of a CMV that introduces a deletion in the target gene is deemed to be the length of the deletion. The mutator region may be a deletion of between one and six
30 nucleobases or between one and three nucleobases. Multiple separated mutations can be introduced by a single CMV, in which case there are multiple mutator regions in the same CMV. Alternatively, multiple CMV can be used simultaneously to introduce multiple genetic changes in a single gene or, alternatively to introduce genetic changes in multiple genes of the

same cell. Herein, the mutator region is also termed the heterologous region. When the different desired sequence is an insertion or deletion, the sequence of both strands have the sequence of the different desired sequence.

5 The 3' terminal nucleobase of each strand may be protected from 3' exonuclease digestion. Such protection can be achieved by several techniques now known to those skilled in the art or by any technique to be developed. For example, protection from 3'-exonuclease digestion may be achieved by linking the 3' most (terminal) nucleobase of one strand with the 5' most (terminal) nucleobase of the alternative strand by a nuclease-resistant covalent linker, such as polyethylene glycol, poly-1,3-propanediol, or poly-1,4-butanediol. The length of
10 various linkers suitable for connecting two hybridized nucleic acid strands is understood by those skilled in the art. A polyethylene glycol linker having from six to three ethylene units and terminal phosphoryl moieties is suitable (Durand et al., *Nucleic Acids Research* 18:6353, 1990; Ma et al., *Nucleic Acids Research* 21:2585-2589, 1993); bis-phosphorylpropyl-trans-4,4'-stilbenedicarboxamide may also be used as a linker (Letsinger et al., *Journal of the*
15 *American Chemical Society* 116:811-812, 1994; Letsinger et al., *Journal of the American Chemical Society* 117:7323-7328, 1995). Such linkers can be inserted into the CMV using conventional solid phase synthesis. Alternatively, the strands of the CMV can be separately synthesized and hybridized, and then forming an interstrand linkage with thiophoryl-containing stilbenedicarboxamide as described in patent application WO 97/05284.

20 The linker can be a single strand oligonucleobase comprised of nuclease-resistant nucleobases (e.g., a 2'-O-methyl, 2'-O-allyl or 2'-F-ribonucleotides). The tetranucleotide sequences TTTT, UUUU and UUCG and the trinucleotide sequences TTT, UUU, or UCG are particularly preferred nucleotide linkers. A linker comprising a tri- or tetra-thymidine oligonucleobase is not comprised of nuclease-resistant nucleobases and such linker does not
25 provide protection from 3' exonuclease digestion.

Alternatively, modification of the 3' terminal nucleobase can protect it from digestion by 3'-exonuclease. If the 3' terminal nucleobase of a strand is a 3' end, then a steric protecting group can be attached by esterification to the 3'-OH, the 2'-OH or to a 2' or 3' phosphate. Suitable protecting groups are 1,2-(ω -amino)-alkyldiols or, alternatively, 1,2-hydroxymethyl-(ω -amino)-alkyls. Modifications that can be made include use of an alkene or branched alkane or
30 alkene, and substitution of the ω -amino or replacement of the ω -amino with an ω -hydroxyl. Other suitable protecting groups include a 3'-methylphosphonate, (Tidd et al., *British Journal of Cancer* 60:343-350, 1989) and 3'-aminohexyl (Gamper et al., *Nucleic Acids Research*

21:145-150, 1993). Alternatively, the 3' or 5' end hydroxyls can be derivatized by conjugation with a substituted phosphorus (e.g., methylphosphonates or phosphorothioates).

Moreover, the 3'-most nucleobase can be made a nuclease-resistant nucleobase to protect the 3'-terminus. Nuclease-resistant nucleobases include PNA nucleobases and 2' substituted ribonucleotides. Suitable substituents include those disclosed in U.S. Patent Nos. 5,334,711; 5,658,731; and 5,731,181 and those disclosed in EP 0 629 387 and EP 0 679 657. The 2' fluoro, chloro, or bromo derivatives of a ribonucleotide or a ribonucleotide having a substituted 2'-O as described in the aforementioned are termed 2'-Substituted Ribonucleotides (e.g., 2'-fluoro, 2'-methoxy, 2'-propyl-oxy, 2'-allyloxy, 2'-hydroxyethyloxy, 2'-methoxy-ethyloxy, 2'-fluoropropoxy, and 2'-trifluoropropoxy substituted ribonucleotides; 2'-fluoro, 2'-methoxy, 2'-methoxyethyloxy, and 2'-allyloxy substituted nucleotides). A "nuclease-resistant ribonucleoside" includes 2'-Substituted Ribonucleotides and all 2'-hydroxyl ribonucleosides other than ribonucleotides (e.g., ribonucleotides linked by non-phosphate or by substituted phosphodiester).

CMV may have a single 3' end and a single 5' end which are the terminal nucleobases of a strand. Alternatively, a strand may be divided into two chains that are linked covalently through the alternative strand but not directly to each other. Where a strand is divided into two chains, the 3' and 5' ends are Watson-Crick base paired to adjacent nucleobases of the alternative strand. In such strands, the 3' and 5' ends are not terminal nucleobases. A 3' end or 5' end that is not the terminal nucleobase of a strand can be optionally substituted with a steric protector from nuclease activity as described above. Alternatively, a terminal nucleobase of a strand is attached to an nucleobase that is not paired to a corresponding nucleobase of the opposite strand and is not a part of an interstrand linker. It has a single "hairpin" conformation with a 3' or 5' overhang. The unpaired nucleobase and other components of the overhang are not regarded as a part of a strand. The overhang may include self-hybridized nucleobases or non-nucleobase moieties (e.g., affinity ligands or labels). In a CMV having a 3' overhang, the strand containing the 5' nucleobase may be composed of deoxy-type nucleobases only, which are paired with ribo-type nucleobase of the opposite strand. For a CMV having a 3' overhang, the sequence of the strand containing the 5' end nucleobase is the different, desired sequence and the sequence of the strand having the overhang is the sequence of the target gene.

The linkage between the nucleobases of the strands of a CMV can be any linkage that is compatible with hybridization of the CMV to its target sequence. Such sequences include the

conventional phosphodiester linkages found in natural nucleic acids. The organic solid phase synthesis of oligonucleobases is described in U.S. Patent No. Re 34,069.

The internucleobase linkages can also be substituted phosphodiester (e.g., phosphorothioates, substituted phosphotriesters). Alternatively, non-phosphate, phosphorus-containing linkages can be used. U.S. Patent No. 5,476,925 describes phosphoramidate linkages. The 3'-phosphoramidate linkage (3'-NP(O)(O)O-5') is well suited for use in CMV because it stabilizes hybridization compared to a 5'-phosphoramidate. Non-phosphate linkages between nucleobases can also be used. U.S. Patent No. 5,489,677 describes internucleobase linkages having adjacent nitrogen and oxygen heteroatoms, and their synthesis. Another preferred linkage is 3'-ON(CH₃)CH₂-5' (methylenemethylimino). Other linkages suitable for use in CMV are described in U.S. Patent No. 5,731,181. Nucleobases that lack a pentosefuranosyl moiety and are linked by peptide bonds can also be used. Such so-called peptide nucleic acids (PNA) are described in U.S. Patent No. 5,539,082; methods for making PNA/nucleotide chimera are described in patent application WO 95/14706. The 2' position end of the internucleobase linkage can be modified (Freier & Altmann, *Nucleic Acids Research* 25:4429-4443, 1997).

Formulation of the Compositions

A polymer (e.g., polyethylene glycol or PEG, polyethylenimine or PEI) can be included in the composition. They could have an average molecular weight of greater than about 500 daltons, preferably greater than between about 10 kd and more preferably about 25 kd (mass average molecular weight determined by light scattering). The upper limit of suitability is determined by the toxicity and solubility of the polymer, but molecular weights greater than about 1.3 Md are possibly less suitable. Alternatively, inert polymeric materials could be formed into nanospheres or microspheres as transfection agents (cf. Leong et al., *Journal of Controlled Release* 53:183-193, 1998; Baranov et al., *Gene Therapy* 6:1406-1414, 1999).

The use of high molecular weight PEI as a carrier to transfect a cell with DNA is described in Boussif et al., *Proceedings of the National Academy of Sciences* 92:7297-7301, 1995. A CMV carrier complex can be formed by mixing an aqueous solution of CMV and a neutral aqueous solution of PEI at a ratio of between about 4 and about 9 PEI nitrogens per CMV phosphate, preferably the ratio is about 6. The complex can be formed, for example, by mixing a 10 mM solution of PEI, at pH 7.0 in 0.15 M NaCl with CMV at a final concentration of between 100 and 500 nM CMV.

A ligand can also be included in the composition. Suitable ligands are those that specifically bind receptors in clathrin-coated pits, transferrin, nicotinic acid, α -bungarotoxin, carnitine, insulin, and insulin like growth factor-1 (IGF-1). In an alternative embodiment, the ligand contain glucosyl moieties, such as glucose. For example, a 1:1 mixture of glucosylated
5 PEI having a ratio of between about 0.4 and about 0.8 glucose moieties per nitrogen and unmodified PEI can be used. The mixture is used in a ratio of between 4 and 9 PEI nitrogens per CMV phosphate, preferably the ratio of CMV phosphate to nitrogen is about 1:6. PEIs having a mass average molecular weight of 25 kd and 800 kd are commercially available from Aldrich Chemical Co., Catalog No. 40,872-7 and 18,197-8, respectively. The optimal ratio of
10 ligand to polyethylene subunit can be determined by fluorescently labeling the CMV and injecting fluorescent CMV/molecular carrier/ligand complexes directly into the tissue of interest and determining the extent of fluorescent uptake according to the method of Kren et al., Hepatology 25:1462-1468, 1997. Furthermore, a basic protein (e.g., histone H1) can be substituted as a polycationic carrier.

15 Transfection agents that at least in part condense the CMV may be used. Alternatively, transfection agents like lipids may form liposomes or other structures that encapsulate the CMV. Many neutral and charged lipids, sterols, and other phospholipids to make lipid carrier vehicles are known.

Synthetic lipids or purified lipid biological preparations, e.g., soybean oil (Sigma) or
20 egg phosphatidyl choline (EPC) (Avanti Polar Lipids) can be used. Other lipids that are useful in the preparation of lipid nanospheres and/or lipid vesicles include neutral lipids, e.g., dioleoyl phosphatidylcholine (DOPC) and dioleoyl phosphatidyl ethanolamine (DOPE); anionic lipids, e.g., dioleoyl phosphatidyl serine (DOPS); and cationic lipids, e.g., dioleoyl trimethyl ammonium propane (DOTAP), dioctadecyldiamidoglycyl spermine (DOGS), dioleoyl
25 trimethyl ammonium (DOTMA), and DOSPER (1,3-di-oleoyloxy-2-(6-carboxy-spermyl)-propyl-amide tetraacetate, commercially available from Boehringer-Mannheim. Additional examples of lipids that can be used in the invention can be found in Gao & Huang (Gene Therapy 2:710-722,1995). Saccharide ligands can be added in the form of saccharide cerebroside, e.g., lactosylcerebroside or galactocerebroside (Avanti Polar Lipids). DPPC
30 (dipalmitoyl phosphatidylcholine) can be incorporated to improve the efficacy and/or stability of delivery. FUGENE 6, LIPOFECTAMINE, LIPOFECTIN, DMRIE-C, TRANSFECTAM, CELLFECTIN, PFX-1, PFX-2, PFX-3, PFX-4, PFX-5, PFX-6, PFX-7, PFX-8, TRANSFAST, TFX-10, TFX-20, TFX-50, and LIPOTAXI lipids are proprietary sources of lipid.

Lipid nanospheres can be constructed by the following process. A solution of phospholipids in organic solvent is added to a small test tube and the solvent removed by a nitrogen stream to leave a lipid film. A lipophilic salt of CMV is formed by mixing an aqueous saline solution of CMV with an ethanolic solution of a cationic lipid. The cationic species can be in about 4 fold molar excess relative to the CMV anions (phosphates). The lipophilic CMV salt solution is added to the lipid film, vortexed gently followed by the addition of an amount of neutral lipid equal in weight to the phospholipids. The concentration of CMV can be up to about 3% (w/w) of the total amount of lipid. After addition of the neutral lipid, the emulsion is sonicated at 4°C for about 1 hour until the formation of a milky suspension with no obvious signs of separation. The suspension is extruded through polycarbonate filters until a final diameter of about 50 nm is achieved. The CMV-carrying lipid nanospheres can then be washed and placed into a pharmaceutically acceptable carrier or tissue culture medium. The capacity of lipid nanospheres is about 2.5 mg CMV/ 500 µl of a nanosphere suspension.

A lipid film is formed by placing a chloroform methanol solution of lipid in a tube and removing the solvent by a nitrogen stream. An aqueous saline solution of CMV is added such that the amount of CMV is between 20% and 50% (w/w) of the amount of lipid, and the amount of aqueous solvent is about 80% (w/w) of the amount of lipid in the final mixture. After gentle vortexing the liposome-containing liquid is forced through successively finer polycarbonate filter membranes until a final diameter of about 50 nm is achieved. The passage through the successively finer polycarbonate filter results in the conversion of polylaminar liposomes into unilaminar liposomes (i.e., lipid vesicles). The lipid nanospheres can then be washed and placed into a pharmaceutically acceptable carrier. About 50% of the added CMV can be entrapped in the vesicle's aqueous core.

A variation of the basic procedure comprises the formation of an aqueous solution containing a PEI/CMV condensate at a ratio of about 4 PEI imines per CMV phosphate. The condensate can be particularly useful when the liposomes are positively charged, i.e., the lipid vesicle contains a concentration of cations of cationic lipids such as DOTAP, DOTMA or DOSPER, greater than the concentration of anions of anionic lipids such as DOPS. The capacity of lipid vesicles is about 150 µg CMV per 500 µl of a lipid vesicle suspension.

Lipid vesicles may contain a mixture of the anionic phospholipid, DOPS, and a neutral lipid such as DOPE or DOPC; negatively charged phospholipids that can be used to make lipid vesicles include dioleoyl phosphatidic acid (DOPA) and dioleoyl phosphatidyl glycerol (DOPG). For example, the neutral lipid may be DOPC and a ratio of DOPS:DOPC between

about 2:1 and about 1:2, preferably about 1:1. The ratio of negatively charged to neutral lipid is preferably greater than about 1:9 because the presence of less than 10% charged lipid results in instability of the lipid vesicles because of vesicle fusion.

5 An optional additive to the composition is an insoluble indicator that will not diffuse a substantial distance in solid tissue from the site of injection. For example, a signal-generating particle mixed into the composition with indicate the injection track. Gene repair and/or a change in physiological resulting from gene repair can then be correlated with localization of the CMV introduced into cells. The signal can be a fluorophore, radioisotope, other emitters of electromagnetic radiation, colloidal metal, contrast agent for ultrasound or electromagnetic
10 radiation, chromagen, or be generated by an enzyme attached to the particle (e.g., alkaline phosphatase, horseradish peroxidase). Similarly, entry into cells can be determined by labeling the CMV, and then visualizing the label or comparing the amount of label in separated extracellular and intracellular fractions. Placement of CMV *in situ* may be guided by soluble or insoluble signals (e.g., fluorophores, radiochemicals, other emitters of electromagnetic
15 radiation, contrast agents) and ultrasonography/radiography, or visualized with fiber optics.

At least some of the CMV and optional agents of the composition may self-assemble upon mixing. They may associate by interactions that are covalent (e.g., linkages with an amino or thiol reactive group, photo adducts) or non-covalent (e.g., hydrogen bonding, electrostatic or hydrophobic forces). The degree of association may be assessed by techniques
20 such as, for example, fluorescence quenching or transfer, light polarization or scattering, electrophoretic mobility, size-exclusion chromatography, and electron microscopy.

Canine Model of Muscular Dystrophy

A composition comprising a CMV packaged in FuGENE™ 6 lipid was introduced into
25 an affected cell and produced dystrophin protein containing exon 7 epitopes. The invention further encompasses the use of alternative lipid carriers that are equivalent to FuGene™ 6 lipid, now known or to be developed. Naked CMV (i.e., introduced into an affected cell without transfection agents like lipids, viral particles, DEAE-dextran, salt and polymeric precipitants, etc.) are not effective in this embodiment of the invention. But it is well within the skill of the
30 art to determine under which circumstances naked CMV could be effectively used for gene repair (e.g., the *mdx* mutation exemplified below).

Correction of the GRMD point mutation, as detected at the mRNA, protein, and genomic DNA levels, was found up to 11 months after a single treatment with the CMV. To

facilitate the analysis of the GRMD model, an exon 7-specific antibody against the portion of the protein deleted by the GRMD mutation provided a unique reagent for discriminating patterns of dystrophin protein expression resulting from successful gene repair to that produced by alternative processing of the mRNA. The critical importance of exon-specific antibodies for unequivocal identification of wild-type dystrophin in muscle fibers has been demonstrated in human myoblast therapy trials. At 11 months post-injection, detectable quantities of normal sized dystrophin were localized in multiple regions within the treated cranial tibialis muscle using the MANEX7B antibody. These results were obtained by both western blot and immunohisto-chemical analyses using the MANEX7B antibody. We estimate that the level of gene repair approaches, but does not exceed, 1% in our studies based on comparative levels of RT/PCR product from the exon 7-deleted mRNA produced by the GRMD allele in the nine-week biopsy sample. To clarify these analyses, RT/PCR primers were specifically selected to discriminate the mutant mRNA and corrected mRNA species from alternately spliced products. Precise quantitative estimates of the level of reversion have proven difficult due to the inherent AT-rich nature of the intron portion of this splice junction, which renders a quantitative method such as allele-specific primer discrimination problematic at best. Thus, we have been limited to qualitative differences rather than quantitative differences between the mRNA/genomic results from the tissues treated with the two CMV used in these experiments versus untreated tissue from the same animal. It is of interest to note that RT/PCR of RNA extracted from the necropsy samples from the right limb treated with the chimera without FuGENE™ 6 lipid failed to produce any detectable exon 7-containing dystrophin mRNA. This is in contrast to the localization seen in both frozen sections taken from the small biopsy sample taken at six weeks for the in situ RT/PCR as well as the immunohistochemistry of the six-week sample. Based on this difference, we believe that the initial frequency of gene repair for the two limbs favored delivery using a carrier vehicle of FuGENE™ 6 lipid for sustained inclusion in nascent dystrophin mRNA of epitope expression of exon 7.

Murine Model of Muscular Dystrophy

The *mdx* mouse strain has a point mutation in the dystrophin gene, the consequence of which is a muscular dystrophy due to deficiency of dystrophin in skeletal muscle. As a test of the feasibility of CMV-mediated gene therapy for muscular dystrophies, a CMV termed MDX1 was designed to induce correction of the point mutation in the dystrophin gene in *mdx* mice. Two weeks after direct injection of MDX1 into muscles of *mdx* mice, dystrophin expression

was detected in clusters of muscle fibers by immunohistochemical analysis. None of these dystrophin-positive fibers were so called "revertant" fibers (which appear spontaneously in *mdx* muscle) as characterized by antibodies directed against the protein products of specific exons of the dystrophin gene. Furthermore, injection of control CMV did not yield any dystrophin-positive fibers. Immunoblot analysis of dystrophin immunoprecipitated from MDX1-injected muscles revealed a single band corresponding to full-length dystrophin. No dystrophin was detected when muscles injected with control CMV were subjected to the same analysis. These results provide the foundation for further studies of CMV-mediated gene therapy as a novel therapeutic approach to muscular dystrophies and other genetic disorders of muscle.

The invention is used to correct a point mutation in the dystrophin gene in the *mdx* mouse. The *mdx* mouse has a point mutation at nucleotide position 3185 in the dystrophin gene that produces a stop codon in exon 23 (Yoon et al., Proceedings of the National Academy of Sciences USA 93:2071-2076, 1996). As a result, there is no dystrophin produced in skeletal muscle of these mice and the muscle fibers undergo necrotic degeneration as in DMD. Direct injection into skeletal muscle of a CMV designed to correct the point mutation resulted in the expression of functional dystrophin in muscle fibers around the site of injection. Lipid was not required in this embodiment of the invention.

EXAMPLES

The following examples are provided for illustrative purposes only and are not to be construed as limiting the invention's scope in any manner.

CANINE MODEL OF MUSCULAR DYSTROPHY

Correction of the GRMD Mutation Requires Lipid

A diagram of the basis of the sequence of the CMV is shown in Figure 2. The CMV is composed of a five-base segment of DNA which defines the complement of the wild-type coding strand sequence at the splice acceptor site of intron 6 of the canine dystrophin gene (Sharp et al., Genomics 13:115-121, 1992) flanked by complementary segments of O-methyl-RNA (10-13 residues), two hairpin turns composed of four dT bases, a 3' GC clamp segment, and a 5' complementary DNA strand which extends across either end of the two O-methyl-RNA segments. The native structure of such a molecule is believed to be a hairpin (Ye et al., Molecular Medicine Today 4:431-437, 1998). Comparison of the nucleotide sequence of the

GRMD mutation and the CMV sequence predicts that the mismatch should be corrected by CMV-mediated gene repair in a treated dog.

An affected male (six weeks of age) from a litter born at the University of Missouri colony was selected for this study. All animals are maintained in the University of Missouri Vivarium according to ACUC and NIH guidelines for the use of animals in research. At 13 months of age, disease symptoms warranted euthanizing the animal. All surgical biopsy and necropsy samples from treated sartorial compartment muscles as well as the left triceps were collected, wrapped in aluminum foil, and snap-frozen in liquid nitrogen. To determine if gene repair mediated by CMV could be used to correct the mutation that causes GRMD, a six week-old affected male was selected for study.

A timeline diagram of the experimental procedures performed on the GRMD affected male is found in Figure 3. At six weeks of age (time point A), CMV designed to correct the GRMD mutation (200 µg from BioSource) was mixed with 200 µg of calf thymus histone H1 (Sigma) and packaged in FuGENE™ 6 lipid plus OPTIMEM media (LTI) in a final volume of 5.0 ml. Proprietary FuGENE™ 6 lipid is commercially available from Roche Diagnostics (<http://biochem.roche.com/techserv/fugene.htm>); it is a proprietary blend of lipids and other components supplied in 80% ethanol, sterile filtered, and packaged in polypropylene tubes. The injectate also contains 7.5 µl/ml of fluorescent microspheres (Molecular Probes) to mark the site of injections.

After surgical exposure of the sartorial compartment, the full 5.0 ml was injected into the cranial tibialis compartment of the left limb using 50 separate injections of 100 µl each. Surgical biopsy samples were taken and snap-frozen in liquid nitrogen at 2 (time point B) and 9 (time point C) weeks post-injection (Bartlett et al., Nature Biology Short Reports 9:163-164, 1998) and at necropsy at 11 months (point E) post-injection from the left leg. A biopsy of control untreated triceps muscle was removed for RNA, western, and immunohistochemical analyses prior to injection.

To determine whether FuGENE™ 6 lipid is required to correct the GRMD mutation, additional CMV from Kimeragen was injected into the contralateral limb during the surgical procedure for the 9 week biopsy (time point C). A biopsy was also taken from the contralateral limb at 6 weeks post-injection (time point D). The protocol for treatment was the same as that used for the left leg with the lone exception that FuGENE™ 6 lipid was not included in the injectate. Force generation studies (diagonal arrows) were performed at the three indicated times. The entire cranial tibialis, long digital extensor and triceps muscles (left and right) were

removed at necropsy at 13 months of age when the animal was euthanized (time point E) due to progressive disease complications.

To summarize the results obtained and further discussed below, we found that a lipid carrier was required for sustained inclusion in nascent dystrophin mRNA of the epitope encoded by exon 7. A composition that did not include FuGENE™ 6 lipid was ineffective and produced no dystrophin protein containing exon 7 epitopes.

RT/PCR Analysis Detects Normal-Sized Dystrophin mRNA in Treated Skeletal Muscle

To investigate whether therapy with CMV that corrected the GRMD mutation would produce a change in the pattern of expression of functional dystrophin in GRMD muscle, total RNA was isolated from frozen sections of biopsy and necropsy material taken at various timepoints after treatment. To isolate RNA, about 10-20 serial frozen 20 µm thick sections from the same segments of muscle used for parallel analyses by western blotting and immunohistochemistry (see below) were made and stored at -80°C in separate RNase-free tubes.

Total RNA was isolated using an RNAEASY kit (Promega). RT/PCR was performed using 5' primer (544) and 3' primers (704 and 120) that bracket exon 7 of the canine dystrophin mRNA (Sharp et al., Genomics 13:115-121, 1992). The primers used in this analysis are shown in Figure 4 positioned relative to the respective sequence. The direction of the arrows indicate 5' primers (right pointing arrows) and 3' primers (left pointing arrows). RT/PCR products were separated on ethidium-stained 1% agarose gels with normal product at 1058 bp and mutant product at 929 bp.

While suggestive levels of normal-sized dystrophin RT/PCR product containing exon 7 were seen in the 2 week sample, the results from the 9 week sample demonstrated that at least as much product from normal-sized mRNA was present in the biopsy as the mutant mRNA. Confirmation of the presence of exon 7 in the PCR product was by sequencing and re-PCR with an exon 7-specific 3' primer and the original 5' primer. Moreover, analysis of a necropsy sample from the left limb (FuGENE™ 6 lipid-treated sample) taken at 11 months reveals that the predominant RT/PCR product was of normal size. Since the level of mutant mRNA is <1% of normal in muscle of GRMD dogs and not visible on a northern blot, we conclude that the frequency of gene repair in these studies produced a similar modest level of normalized dystrophin mRNA.

Gene Repair of GRMD Mutation Corrects Endogenous Sequence in Affected Tissue

To confirm that the invention had actually corrected the mutation, genomic DNA was isolated from additional serial frozen sections and its nucleotide sequence was determined. Genomic DNA was isolated from twenty 20 µm frozen sections from untreated tricep muscle, treated cranial tibialis (CT), and normal CT muscles using a commercial kit from Qiagen. PCR of genomic DNA was performed using intronic primers that bracket exon 7 in the canine dystrophin gene (Bartlett et al., American Journal of Veterinary Research 57:650-654, 1996).

The GRMD mutation produces a novel Sau96 recognition site such that digestion of the 310 bp genomic PCR product is diagnostic of the mutant allele. Thus, to enrich for corrected GRMD sequences that could be detected by PCR amplification, all samples were digested with Sau96 to deplete GRMD alleles that had not undergone gene repair: reactions were stopped after 10 cycles of PCR with bracketing primers, submitted to digestion with Sau96, extracted with phenol/chloroform and precipitated from ethanol. The Sau96-digested samples were amplified for another 25 cycles and 310 bp bands from each were separately ligated into the TA cloning vector pCR1 (Invitrogen). After analytical digestion with Sau96, all clones from the untreated triceps muscle cut to completion which is indicative of the GRMD allele and all clones from normal CT muscle were resistant to digestion. Clones were sequenced using an Applied BioSystems automated sequencer at the University of Miami Cancer Center DNA Core. Examination of 50 clones from the left CT muscle identified three that demonstrated a pattern of digestion with Sau96 indistinguishable from that obtained from a canine muscle sample for the normal allele. These three clones were sequenced and each contained the corrected sequence containing the functional splice acceptor site.

In no case were we able to detect a two-base change in these clones of PCR products. This may be due to a bias imparted by the analysis of only Sau96 resistant clones or to a lower efficiency of changing two bases as opposed to one. It was also possible that cloning with this technique may have selected for only single-base changes due to the inclusion of only the 3' base change within the Sau96 recognition site. Screening a larger number of clones (e.g., 600 to 1000) by sequencing might have detected a 5' base change.

Quantitation of Gene Repair Events by RT/PCR

Accurate quantitative estimates of CMV-mediated changes in endogenous sequences have proven difficult due to the inherent AT-rich nature of the intron portion of this splice junction, which renders a quantitative method such as allele-specific primer discrimination

problematic at best. We have used quantitative RT/PCR to demonstrate that inclusion of exon 7 in the dystrophin mRNA from treated GRMD muscle exceeded 10% of normal levels in an isolated sample.

5 Total RNA were extracted from frozen sections collected separately from tissue
harvested at biopsy or necropsy and stored frozen at -80°C. Control and experimental muscle
tissue sections were extracted using the TRIAGENT RNA isolation kit (Molecular Research
Center). RNA concentrations were determined by spectrophotometry and their integrity was
verified by electrophoretic analysis. The RT/PCR reaction was performed according to the
manufacturer specifications using the *C. therm* RT/PCR kit (Roche) and sequence-specific
10 RT/PCR primers which bracketed the GRMD mutation, a deletion of exon 7 from the mRNA
due to a point mutation in the consensus splice acceptor site of intron 6 (Sharp et al., Genomics
13:115-121, 1992). Primer 278 (canine dystrophin forward) was from exon 1 beginning with
the start codon, 5'-ATGCTTTGGTGGGAAGAAGTAGAG-3' (SEQ ID NO:9) and primer 120
(canine dystrophin reverse) was from exon 8 at positions 990-967 in the cDNA, 5'-
15 GTCACCTTTAGGTGGCCTTGGCAAC-3' (SEQ ID NO:10).

Nested canine-specific primers located at 538-568 bp spanning the exon 5/6 junction
and at 874-846 spanning the exon 7/8 junction were used to specifically amplify the normal
canine cDNA in the dilution series. The forward primer spanning the exon 5/6 junction was 5'-
GATTTGGAATATAATCCTCCA(TGGCAGGTC-3' (SEQ ID NO:13) and the reverse primer
20 spanning the exon 7/8 junction was 5'-AGTGGTGGCAACATCTTCAGGATCAA-3' (SEQ ID
NO:14). The sequence of all canine dystrophin primers was determined by sequencing two
clones obtained from RT/PCR of normal canine skeletal muscle RNA using dystrophin primers
based on the human cDNA beginning at the first base (5' primer) and ending at position 1505
bp (3' primer).

25 To insure constant input mRNA from each sample, all total RNA samples were first
normalized using this primer set for the housekeeping gene GAPDH with the forward primer
5'-ATGATGACATCAAGAAGGTGGTGAAGC-3' (SEQ ID NO:11) and the reverse primer
5'-TCTCTCCTCCTCGCGTGCTCTTGCTG-3' (SEQ ID NO:12). GAPDH gene transcripts
were amplified using parallel RT/PCR reactions with a constant sample volume (2 µl) and
30 quantitated using a standard curve generated from normal muscle total RNA. Thereafter, all
total RNA input values for dystrophin RT/PCR reactions were normalized to the values
generated for GAPDH quantitation. All values for GAPDH production fell within a 2-fold
range thus minimizing the range of input template volumes for all dystrophin quantitations.

The GAPDH and dystrophin products were not co-amplified in the RT/PCR reactions because several artifactual bands were produced by the presence of both sets of primers which prevented quantitation of either in the LIGHTCYCLER thermal cycler (Roche).

RT/PCR was performed using the following program in a Perkin Elmer 480 PCR
5 machine: cDNA synthesis (30 min at 53°C and denaturation at 95°C for 5 min), then 20 cycles of PCR amplification using: denaturation at 95°C for 30 sec, annealing at 56°C for 45 sec, and extension at 68°C for 60 sec. A final polishing step of 72°C for 7 min was performed. A nested 5' primer located at 538-568 bp spanning the exon 5/6 junction and primer 120 were combined to re-amplify the various dystrophin RT/PCR products to confirm predicted sizes of
10 a 452 bp product from normal dys mRNA and 333 bp product from GRMD mRNA reflecting the deletion of exon 7. The products were submitted to a melting curve analysis using the LIGHTCYCLER thermal cycler, and a 3°C difference was noted (i.e., GRMD T_m = 81°C, Normal T_m = 84°C).

Alternatively, real-time PCR amplification was performed with a fluorescent LIGHT
15 CYCLER thermal cycler equipped with software that follows the PCR reaction "on-line" step-by-step through all the phases. It also provides us with melting curve analysis and calculations of melting temperature [T_m] of the PCR product. Moreover, quantitation of experimental samples is provided when a standard concentration curve is included in the assay. RT/PCR product from normal muscle was used to generate a standard concentration curve beginning
20 with a 1:10 dilution (0.1X) and through successive 1:10 dilutions down to $1:10^5$ (0.00001X).

Quantitative PCR was performed in a LIGHTCYCLER thermal cycler in a final volume of 20 μ l containing 2 μ l of ready-to-use reaction mix 10 (X). DNA Master SYBR Green I (Roche) was preincubated 5 min at room temperature with 0.55 μ g of TAQSTART antibody (Clontech), 3 mM $MgCl_2$, 0.5 μ M of each primer, and 2 μ l of either the RT/PCR dilution series
25 or a 1:200 dilution of the experimental RT/PCR sample as template. The program to amplify exon-specific products used an initial denaturation step of 95°C for 20 sec to inactivate the Taq antibody; 65 cycles of denaturation at 96°C for 5 sec/annealing at 63°C for 4 sec/extension at 72°C for 30 sec; and acquisition of fluorescence for all samples after heating to 82°C. Thus, the fluorescence is acquired above the T_m of the mutant product (81°C) to insure that the
30 normal product in all samples is measured by fluorescence quantitation. The expected size for the normal dystrophin amplification product is 334 bp.

After each quantitative PCR was completed, a melting curve analysis was performed by heating to 96°C, and cooling to 35°C at 20°C/sec followed by heating to 96°C at much slower

rate (0.1°C/s) and acquiring fluorescence continuously. Identity of PCR products was verified by melting temperature [T_m] and electrophoresis on 1% agarose gel.

All measurements were taken at 83°C to measure the exon 7/8-specific product. The cycle number in which fluorescent signal begins to accumulate is inversely proportional to the starting concentration of exon 7-containing dystrophin cDNA in the template sample. PCR products produced in the indicated samples all contain the expected 330 bp product when run on a 2% NUSIEVE agarose (FMC) gel and stained with ethidium bromide. Since all reactions are taken to equilibrium (completion) during the course of the real-time PCR, the standard curves do not reflect a gradient of concentration when run on this gel. Of critical importance to note, the sample from the affected, untreated tissue, RTCT 2 weeks, contained no product. Moreover, the excellent agreement of the concentration estimates for the standard curves with the expected values, and the production of the appropriately-sized products, demonstrates that the 3' primer used to detect the exon 7/8 junction in the cDNA from the original RT/PCR products provided the appropriate specificity for detecting the presence of inclusion of exon 7 in the cDNA from the experimental samples.

In Situ RT/PCR of Treated Skeletal Muscle Localizes Gene Repair Events

To determine what the pattern of distribution of gene repair was in the injected muscle, we performed *in situ* RT/PCR on frozen sections from normal, GRMD muscle, and the 6 week injected sample from the right leg. Frozen sections of muscle from normal, GRMD mutant, and GRMD injected muscle were prepared on SUPERFROST slides (Fisher Scientific) using a Leica 3000 cryomicrotome. After overnight fixation in 10% buffered formalin, slides were rinsed twice in fresh PBS, then digested for 17 min in 5 µg/ml pepsin. This permitted infusion of RT/PCR reagents into the extensively fixed tissue. Slides were then rinsed in two changes of fresh PBS, treated with RNase free-DNAase to remove nuclear DNA as template from the subsequent RT/PCR reaction, and finally rinsed in four changes of fresh PBS. The slides were covered using *in situ* chambers (RPI, Sci.). Using RT/PCR 3' primers from within exon 7 (459 and M23) and a 5' primer which spans intron 6 (354) in genomic DNA (begins in exon 5 and ends in exon 6), RT/PCR was performed using the Roche/Boehringer Mannheim single tube TITAN RT/PCR kit (i.e., a master mix containing the single enzyme TthI for performing both RT and PCR in a single tube) in the presence of dATP-biotin to label all PCR products with biotin.

Streptavidin conjugated with alkaline phosphatase (AP) and ELF-97 fluorochrome (Molecular Probes) were used to localize the biotinylated PCR products. Thus, after RT/PCR, slides were rinsed twice with fresh PBS and then treated at room temperature with streptavidin alkaline phosphatase derivative to bind the in situ biotinylated PCR product. Then the slides were again rinsed with three changes of fresh PBS, followed by 5 min exposure to the ELF-97 fluorochrome according to the supplier's instructions. ELF-97 fluorochrome is a soluble, pale blue fluorescing phosphate in its original form but upon cleavage by AP, a precipitate is produced that is brightly yellow-green in fluorescence at the sites of biotin incorporated into PCR product. A DAPI long-pass filter (Leitz) was used to visualize this signal from biotin.

Examination of negative control sections from GRMD triceps muscle obtained via biopsy prior to injection revealed complete absence of exon 7 across the entire section. But positive control sections from normal canine muscle expressed exon 7 across the entire section. Experimental sections from injected GRMD muscle had modest localization of exon 7 across the entire section particularly near to fluorescent microspheres indicating proximity to sites of injection. At high magnification, the injected samples show discrete localization of exon 7-containing dystrophin mRNA at the periphery of fibers where one would expect the myonuclei to be located. These results suggest that modest reversion occurred in multiple nuclei proximal to the sites of injection.

Preparation of Exon 7-Specific Monoclonal Antibodies

Frozen sections of 6 μm of thickness from untreated tricep muscle, injected cranial tibialis (CT) muscle, and normal CT muscle were made using a Leica 3000 cryomicrotome and applied to SUPERFROST slides. Primary monoclonal antibodies against dystrophin included a commercially available antibody specific for the carboxy-terminal region (Novacastra) or exon 7-specific as described below. Primary antibody was applied directly to slides at 1:20 dilution in the presence of 5% normal goat serum, while a goat anti-mouse secondary antibody labeled with FITC (Sigma or Jackson Immunesiences) was used to provide a fluorochrome for localization of dystrophin. Slides were counter-stained for 10 min with DAPI (Sigma) at 15 $\mu\text{g/ml}$. Images were captured using 1/8 sec pixel accumulation as TIFF files with an Optonics cooled CCD camera and ScionImage frame-grabber installed in a PowerMac G3 and converted to Photoshop JPEG files for printing on an HP 5M Color Laserprinter.

Initial western blotting and histochemical analysis of the 2 and 9 week samples obtained from tissue of the left limb, as well as the 6 week sample taken from the right limb

using a commercially available carboxy-terminal dystrophin antibody (Novacastra), suggested no detectable increase in dystrophin protein and modest evidence of dystrophin positive fibers located in the region of the injection site marked by fluorescent microspheres. But the levels were no higher than background when compared to uninjected sample from the triceps muscle taken from the same animal prior to therapy. To increase specificity in the immunological analyses, an exon 7-specific antibody was generated for use.

Dystrophin cDNA (cf27 in pUC plasmid from Prof. Kay Davies) was digested with BamHI and NcoI. The 1640 bp fragment from exon 4 to exon 16 was purified and ligated into pMW172 cut with the same restriction enzymes. After electroporation into *E. coli* BL21(DE3), protein expression was induced by 0.4 mM IPTG for 3 hr. Inclusion bodies were isolated by sonication and extracted sequentially with increasing concentrations of urea (2M, 4M, 6M and 8M in PBS). A 5 µg/ml solution of recombinant protein in 8M urea was used to immunize BALB/c mice and monoclonal antibodies were produced by the hybridoma fusion method. Supernatants were screened by ELISA with recombinant proteins and positive wells (110 out of 288) were further tested for reaction with both native dystrophin (immunolocalization at muscle membrane) and denatured dystrophin (binding to an about 427 kd band on western blots of human muscle proteins). Fourteen wells that passed this screening process were cloned twice by limiting dilution to establish the hybridoma lines. Ig subclass was determined using a mouse isotyping kit (Serotec). Control blots with normal human lung showed that only one mouse mAb (MANEX1011E) cross-reacted with utrophin.

Fourteen mAbs raised against a fragment of dystrophin encoded by exons 4-16 were mapped by western blotting with fragments produced by PCR. Exon 7-specific mAbs, for example, recognize an exon 7-16 fragment, but do not recognize exon 8-16 or any smaller fragment. This shows that exon 7 is essential for binding, and we may be confident that the exon 7-specific mAbs will not recognize "revertant" dystrophins lacking exon 7.

Subconstructs of the pMW172:exon 4-16 construct were produced by PCR for epitope mapping. Forward primers with added BamHI sites were synthesized by the Human Genome Mapping Resource Center (Cambridge, UK) as follows: exon 6 (ctcggatcccagggtcaaaaatgtaag, SEQ ID NO:15), exon 7 (ggggatccaggccagacctatttgac, SEQ ID NO:16), exon 8 (ggggatccgatgttgataccacctatc, SEQ ID NO:17), exon 10 (ggggatcccatttgaagctcctga, SEQ ID NO:18) and exon 12 (ggggatcccatagagttttaatggatctc, SEQ ID NO:19). The reverse primer in the pMW172 sequence was gttattgctcagcggtggcagcag (SEQ ID NO:20). PCR products were digested with

BamHI and EcoRI and cloned into pMW172 digested with the same enzymes. Each mAb was tested for binding to the expressed proteins on western blots.

Mixtures of recombinant protein fragments of dystrophin corresponding to exons 6-16, 8-16, 4-16, 7-16, 10-16, and 12-16 were loaded as a strip onto 12% acrylamide gels and separated by SDS-PAGE. Along with the expected dystrophin fragments, degradation products were also present. After electroblotting, monoclonal antibodies were tested on each blot using a miniblotted apparatus as described by Thanh et al. (American Journal of Human Genetics 56:725-731, 1995). The 14 mAbs that were analyzed are shown in Table 1. MANEX1216E does not react with the smallest degradation product and hence recognizes a different epitope from 1216A-D. It is also the only MANEX1216 mAb to recognize native dystrophin in muscle sections. The MANEX7B mAb was selected for further analyses due to strong reactivity to exon 7 and minor reactivity to exon 8.

TABLE 1. Characterization of 14 monoclonal antibodies produced from a dystrophin fragment encoded by exons 4-16.

Name	Clone Number	Ig Class	Exon Mapping	IMF	Blot
MANEX6	4H4	G1	6	Weak	weak
MANEX7A	5D12	G1	7	weak	weak
MANEX7B	8E11	G1	7	+	+
MANEX7C	6F7	G1	7	+	+
MANEX1011A	8A12	G1	10-11	+	+
MANEX1011B	1C7	G2a	10-11	+	+
MANEX1011C	4F9	G1	10-11	+	+
MANEX1011D	7G5	G1	10-11	+	+
MANEX1011E	8H7	G2a	10-11	+	+
MANEX1216A	5A4	G2a	12-16	weak	+
MANEX1216B	6B11	G2a	12-16	weak	+
MANEX1216C	8C8	G1	12-16	weak	+
MANEX1216D	8D11	G1	12-16	weak	+
MANEX1216E	2G10	G1	12-16	+	+

mAbs were tested for binding to native dystrophin by immunofluorescence microscopy (IMF) of human muscle sections and for binding to denatured dystrophin as determined by separate Western blot (Blot) of human muscle extract. Although two mAbs were "weak" on human muscle blots, they reacted strongly on blots of recombinant protein.

Detection of Exon 7-Epitope by Western Blotting After Gene Repair

Western blotting of lysed GRMD skeletal muscle was performed according to Arahata et al. (Proceedings of the National Academy of Sciences USA 86:7154-7158, 1989). Ten to 20 frozen sections were collected from untreated triceps muscle, right cranial tibialis (CT) muscle, right long digital extensor (LDE) muscle, left LDE muscle, CT muscle from a normal dog, and left CT muscle. Cryomicrotome sections of 20 μ m thickness from samples of various types of canine muscle samples were separately collected and stored at -80°C until gels were prepared for electrophoresis. Care was taken to be certain that fresh blades were used after positive control samples were sectioned.

Tissue sections were lysed in buffer (1% SDS, 10 mM EDTA, Tris pH 8.0, and 50 mM DTT), boiled for 3 min, then cleared by centrifugation at 14,000 rpm in a microfuge for 5 min. Samples (3-10 μ l) were loaded onto 3.5-12% laemmli gradient gels with 3% stacking gels and separated in a constant voltage electric field of 60 V per cm for 16 hr. Electroblothing was in transfer buffer (20% methanol, Tris glycine) onto nitrocellulose (Amersham) for 3 hr in a Hoeffer TRANSBLOT electrophoresis chamber. A 1:100 dilution in TBST of the primary antibodies in Table 1 (e.g., exon 7-specific antibody MANEX7B) was incubated for 60 min with the transferred membrane. The membrane was washed extensively and probed with an IMMUNESTAR chemoluminescent kit (goat anti-mouse, BioRad) to detect the MANEX7B mAb bound to the membrane. Kodak XL-R film was exposed for 15 sec, and then processed using a UMAX POWERLOOK II scanner and Photoshop LE computer program. Results were stored on a UMAX Mac-compatible computer.

To investigate whether increases in RT/PCR product containing exon 7 correlated with restoration of normal dystrophin, western blot analyses were performed using the MANEX7B mAb. When samples taken at necropsy were studied using this antibody, restoration of normal sized dystrophin protein containing exon 7 epitope was observed. This is indicative that the treatment with chimera produced a modest level of gene repair detectable at 11 months post injection. While both the left cranial tibialis (CT) muscle, in particular, and the long-digital extensor (LDE) muscle, to a lesser extent, revealed the expected high molecular weight band co-migrating with the normal muscle sample, no significant high molecular weight of dystrophin protein containing exon 7 epitopes was found in the right limb at necropsy. As expected, no high molecular weight protein was found in untreated GRMD muscle samples. Due to limitation of sample size, no samples from the 2, 6 or 9 week timepoints could be included in these analyses. But expression of a normal-sized dystrophin protein containing an

epitope encoded by exon 7 was found 11 months after CMV treatment, and provided evidence that modest levels of gene repair of the GRMD mutation had occurred in the left leg.

Detection of Exon 7-Epitope by Fluorescent Immunohistochemistry After Gene Repair

5 To determine the pattern of dystrophin distribution in the treated skeletal muscle, an epitope encoded by exon 7 was localized on frozen sections taken at necropsy. Frozen sections were blocked with normal goat serum, incubated with MANEX7B mAb as primary antibody and goat anti-mouse FITC-conjugated secondary antibody (Sigma), and counter-stained with DAPI (15 µg/ml). MANEX7B mAb was localized using an FITC fluorescence bandpass filter
10 while cells were visualized using a triple bandpass filter for DAPI fluorescence. Specificity of the MANEX7B mAb was confirmed by finding that it did not localize to untreated GRMD triceps muscle. In contrast, peripheral staining of a small percentage of fibers was observed in the sections taken from both the right and left cranial tibialis (CT) muscles, while the positive control muscles demonstrated a pattern of normal CT muscle staining of wild-type dystrophin.

15 As each injected muscle received numerous injections, positive fibers were found in clusters proximal to the injection track and usually were no more than about 2-3 mm from an injection site. Due to limiting sample mass, biopsy samples from the 2 and 9 week were not tested. Interestingly, no exon 7-epitope was found in the right CT muscle at necropsy. But the localization of the exon 7-epitope to the periphery of muscle fibers 11 months after treatment
20 of the left CT muscle further confirms that gene repair of the GRMD mutation has occurred after treatment. The difference between the two treatments was the use of FuGENE™ 6 lipid as a carrier in the left limb. Based on similar results from parallel studies reported previously in the *mdx* mouse, we suggest that the chimera was more readily introduced into myonuclei using the FuGENE™ 6 lipid carrier, and thus was able to sustain higher levels of long-term
25 expression of functional dystrophin.

Discussion of Results

In a canine model of Duchenne muscular dystrophy (GRMD), a point mutation within the splice acceptor site of intron 6 leads to deletion of exon 7 from the dystrophin mRNA and
30 the consequent frameshift causes early termination of translation. A hairpin-shaped DNA and RNA chimeric oligonucleobase (i.e., a chimeric mutational vector) was designed to correct the chromosomal mutation to wild-type, possibly by inducing the cell's mismatch repair mechanism. Correction of this point mutation allows appropriate splicing of the dystrophin

transcript to include exon 7. Direct injection of the CMV into the skeletal muscle of the cranial tibialis (CT) compartment of a six-week old affected male dog, and subsequent analysis of biopsy and necropsy samples, demonstrated *in vivo* reversion of the GRMD mutation which was sustained for 11 months. RT/PCR analysis of exons 5-10 demonstrated increasing levels of exon 7 inclusion with time. An exon 7-specific dystrophin antibody confirmed synthesis of normal-sized dystrophin product and positive localization to the sarcolemma. Chromosomal reversion in muscle tissue was confirmed by RFLP/PCR and sequencing the PCR product. This is the first long-term demonstration of reversion of a point mutation in muscle of a live animal using a CMV. *In vivo* delivery of a CMV and lipid composition provides an alternative to myoblast transplantation or viral gene therapy for the treatment of Duchenne dystrophy and other muscular dystrophies that addresses deficiencies of such methods.

Since the CMV used above actually modifies the mutant gene while maintaining all of the native control elements for dystrophin expression, production of dystrophin from a threshold level of corrected genes would be predicted to permit normalization of dystrophin expression patterns in the skeletal muscle. Expanded studies with multiple animals would also permit force generation analyses to correlate potential strength improvement produced from expression of normalizes dystrophin. Moreover, as the resulting dystrophin gene expression patterns reported here are subclinical, methods to improve the frequency of reversion are under consideration. These improvements would include: 1) higher concentrations of CMV delivered either as a single bolus or in serial administrations, 2) extended delivery via an implantable osmotic pump, 3) addition of carrier molecules such as modified polyethyleneimine (PEI) or ligands targeting skeletal muscle cells, and 4) alternate methods of physical introduction such as electroporation.

Based on a previous report in liver using a chimera to mutate the factor IX gene in rats, higher levels of gene modification were achievable by improving delivery of CMV. A putative clinically-relevant threshold of dystrophin expression to prevent the dystrophic phenotype has been suggested to be 20% of normal levels. Thus, strategies which produce higher levels of reversion may be useful since CMV have little inherent capacity for inducing an immune response. As reported previously for liver, serial administration of CMV in dystrophic muscle might have additive effects and may result in achievement of clinically relevant levels of gene modification which would be measurable by force-generation in this animal model for Duchenne muscular dystrophy.

Furthermore, we believe the GRMD model should also be useful for analyzing the potential of using CMV for restoration of reading frame caused by deletions. The fact that exon 7 is missing from the dystrophin mRNA in dogs with this mutation actually simulates an exon 7 genomic deletion. Thus, a CMV designed to restore reading frame by modifying the coding sequence beginning in exon 8 to match the reading frame from exon 6 would be predicted to produce a protein that would be Becker-like and may have sufficient function to normalize the muscle in this model.

MURINE MODEL OF MUSCULAR DYSTROPHY

Design and Synthesis of Chimeric Mutational Vector

The primary sequence of the CMV, termed MDX1, was designed to correct the point mutation in the *mdx* dystrophin gene (Figure 5). Two CMV were used as controls with identical results: one has a sequence homologous to a region of the dog dystrophin gene (a 28-bp region spanning intron 6 and exon 7) and the other was used to the sickle-cell mutation in a globin gene (designated SC1; Cole-Strauss et al., Science 273:1386-1389, 1996). The flanking sequences for both were the same as the flanking sequences in MDX1.

CMV were synthesized as previously described (Sicinski et al., Science 244:1578-1580, 1989). Oligonucleobases were prepared with DNA and 2'-O-methyl RNA phosphoramidite nucleoside monomers on a Perseptive Biosystems Expedite Nucleic Acid Synthesizer, purified by HPLC and quantified by UV absorbance. The Cy3-MDX1 CMV were purified using ABI OPC reverse phase purification cartridges and ethanol precipitated twice. More than 95% of the purified oligonucleobases were determined to be of full length.

Direct Injection of CMV for Gene Repair

Mice of the *mdx* strain (C57BL/10ScSn-*mdx*) were obtained from Jackson Lab (Bar Harbor, ME) and were handled in accordance with guidelines of the Administrative Panel on Laboratory Animal Care of Stanford University. Mice were anesthetized with a ketamine/xylazine cocktail (doses: 125 mg/kg ketamine; 25 mg/kg xylazine). For each injection, the skin over the tibialis anterior muscle was shaved, sterilized, and incised. CMV was dissolved in PBS at a concentration of 4 mg/ml, and the solution was drawn up into a 10 µl Hamilton syringe with a 30 gauge needle. The needle was inserted along the rostro-caudal axis of the muscle into the center of the muscle belly, and 20 µg of the CMV solution was injected in a volume of 5 µl. After the injection, the skin was sutured closed.

Histologic and Fluorescent Immunohistochemical Analyses

Mice were sacrificed at different times after CMV injection, and the tibialis anterior muscles were dissected. The muscles were embedded in OCT mounting compound (Miles),
5 frozen in isopentane cooled in liquid nitrogen, and stored at -80°C. Frozen sections were collected on gelatin-coated slides and stored at -20°C. Serial cross-sections (7 µm thick) were collected along the entire length of the muscle at intervals of 200-300 µm.

Alternatively, for analysis of Cy3 fluorescence after injection of Cy3-MDX1 CMV, muscle sections were warmed to room temperature, hydrated in PBS for 5 min, and cover-
10 slipped using an aqueous mounting medium. Sections were examined using a Zeiss Axioskop fluorescent microscope.

For dystrophin immunohistochemical staining, an antibody directed against the rod domain of the dystrophin protein (MANDYS-8; Sigma) was used at a dilution of 1:400. Specific antibody binding was detected with an Alexa-coupled, goat-anti-mouse secondary
15 antibody (Molecular Probes) at a dilution of 1:1000. Controls for specific staining included sections treated with no primary antibody. The number of dystrophin-positive fibers in a given muscle was determined in the serial section containing the greatest number of fibers. To test for revertant fibers, an antibody directed against the protein product encoded by exon 26 of the dystrophin gene (MANDYS-18; a gift from Dr. Glenn Morris) was used at a dilution of 1:3 in
20 place of the MANDYS-8 antibody.

For routine histological analysis, sections adjacent to those processed for fluorescence microscopy were stained with hematoxylin and eosin (H&E). The needle track was easily identified in H&E-stained sections both by the characteristic changes in muscle architecture created by the needle injury and by the reproducible location in the muscle. Furthermore, in
25 muscles injected with Cy3-MDX1, the distribution of the fluorochrome corresponded exactly with the location of the needle track identified in H&E-stained adjacent sections.

Immunoprecipitation and Immunoblot Analyses

For immunoblot analysis, muscles were dissected and homogenized in RIPA buffer
30 consisting of 150 mM NaCl, 50 mM Tris-HCl, 5 mM EDTA, 5 mM EGTA, 0.5% deoxycholate, 1% NP40, 20 µg/ml leupeptin, 20 µg/ml aprotinin, 100 µg/ml PMSF, and 50 mM DTT. For each sample, the protein concentration was determined using the Bio-Rad protein assay. When dystrophin was immunoprecipitated prior to electrophoresis, equal

amounts of protein (6 mg) from precleared extract were immunoprecipitated using the MANDYS-8 anti-dystrophin antibody (1:100) for 3 hr on ice, followed by protein-G-agarose for 1 hour. Samples were run on 5% SDS-polyacrylamide gels, transferred to 0.45 μ m nitrocellulose membranes (Schleicher and Schuell), and probed with mouse monoclonal antibodies to dystrophin (MANDYS-8, 1:400 dilution, or MANDYS-18, 1:100 dilution) followed by a horseradish peroxidase-coupled sheep-anti-mouse secondary antibody. Specific antibody binding was detected by an enhanced chemiluminescence system (Amersham).

Distribution of Injected CMV

In order to assess first the uptake and distribution of the CMV after injection, fluorochrome-coupled MDX1 was injected into the tibialis anterior muscles of *mdx* mice. The distribution of the fluorescent label was examined in muscle sections at different times after injection and was very characteristic. Labeled fibers were seen in two contiguous areas - a linear pattern defining the track of the needle and a cluster at the end of the needle track at the actual injection site. This pattern was clearly discernible 4 hr after injection and persisted with little apparent change over the next 24 hr. By 48 hr after injection, the intensity of the fluorescent signal was greatly diminished, and it was barely detectable 72 hr after injection. Presumably, this decline in signal represents the metabolism of the CMV and provides some evidence of the stability of these molecules in the cell.

Dystrophin Expression in MDX1-Injected Muscles

To test the efficacy of MDX1 to effect gene repair in *mdx* mouse muscle, tissue sections were examined for dystrophin expression two weeks after MDX1 injections. Expression was seen only along the needle track and at the injection site. Dystrophin immunohistochemical staining around the injection site in two muscles injected with MDX1 was also examined. In each muscle, dystrophin-positive fibers were detected in a pattern similar to the pattern of fluorescent label seen with the fluorochrome-labeled CMV, either along a linear track or in a small cluster. When control CMV were injected, no dystrophin-positive fibers were detected in the vicinity of the injection site.

In order to obtain a quantitative measure of the efficacy of this procedure, the number of dystrophin positive fibers was counted two weeks after a single MDX1 injection in a series of muscles. The number of dystrophin-positive fibers ranged from a low of nine to a high of 32 in these muscles. These numbers represent a range of about 10-20% of the number of fibers

brightly stained by fluorescent CMV 24 hr after injection. Thus, only a subset of fibers that took up the CMV produced sufficient dystrophin to be detected by immunohistochemical staining.

5

Detection of Revertant Fibers

In *mdx* mouse muscle as well as in human muscle from patients with DMD, there is an increase in the appearance of dystrophin-positive fibers, so-called 'revertant' fibers, with age (Hoffman et al., Journal of Neurological Sciences 99:9-25, 1990). For *mdx* muscle, the molecular basis of this reversion has been postulated to be spontaneous, somatic mutations resulting in either in-frame deletions around and including exon 23 (which contains the point mutation), or alternative splicing reactions which would produce transcripts that excluded exon 23. This hypothesis is supported by analysis of revertant fibers with exon-specific antibodies to dystrophin and by nested PCR analysis of transcripts in *mdx* and DMD muscle (Wilton et al., Muscle and Nerve 20:728-734, 1997; Thanh et al., American Journal of Human Genetics 15 56:725-731, 1995).

The negative results with the control CMV argue against any non-specific (i.e., sequence-independent) effect of the experimental procedures leading to an increase in the number of revertant fibers as an explanation for the dystrophin-positive fibers seen after MDX1 injection. Still, to rule out this possibility with greater certainty, antibodies directed against the protein products of exons that are rarely, if ever, expressed in revertant fibers (generally, exons 20-30) were used (Wilton et al., Muscle and Nerve 20:728-734, 1997; Lu & Partridge, Journal of Histochemistry and Cytochemistry 46:977-983, 1998).

A monoclonal antibody directed against exon 26 stained the same fibers as those detected with the antibody directed against a distant region of the dystrophin protein, providing further evidence that dystrophin expression in MDX1-injected muscles was not due to an increased generation of revertant fibers. When the exon 26-specific antibody was used to stain the rare dystrophin-positive fibers away from the site of injection, the staining was negative as would be expected for a revertant fiber (Wilton et al., Muscle and Nerve 20:728-734, 1997; Lu & Partridge, Journal of Histochemistry and Cytochemistry 46:977-983, 1998).

30 As a further demonstration that the dystrophin immunoreactivity found in MDX1-injected muscle represented a correction of the point mutation and thus the expression of full-length dystrophin, the muscles were examined for dystrophin expression by immunoblot analysis. Because of the low number of dystrophin-positive fibers seen in muscle sections,

dystrophin expression was undetectable by standard Western blot analysis. This was not surprising since the percentage of dystrophin-positive fibers generated from MDX1 injections in any given muscle was, at best, approximately 1-2% of the total number of fibers. Therefore, an anti-dystrophin antibody was used to immunoprecipitate any dystrophin that might be present, and the immunoprecipitate was then subjected to immunoblot analysis. Using this approach, a single band was detected at a molecular weight corresponding to full-length dystrophin (427 kd) in MDX1-injected muscles. In muscles injected with control CMV, no such band was detected. That MDX1 is inducing single-base exchange, thus correcting the *mdx* mutation, is supported by the finding of full-length dystrophin by immunoblot analysis. The generation of revertants by somatic deletions or alternative splicing would be expected to produce truncated forms of the protein.

CMV are taken up into mature myofibers as evidenced by the appearance of fluorescent label in myofibers within 4 hr of injection of fluorescently labeled compounds. Expression of dystrophin in mature fibers within two weeks of injection of MDX1 chimeric mutational vector suggests that CMV-induced gene correction may occur in post-mitotic cells. However, it is also possible that the gene correction event could have occurred in proliferating myoblasts which subsequently fused with the mature fibers. Experiments are ongoing to test this possibility by injuring muscle to stimulate myoblast proliferation prior to CMV injection.

Results confirming the above are published as Rando et al., Proceedings of the National Academy of Sciences USA 97:5363-5368, 2000; Bartlett et al., Nature Biotechnology, in the press, June 2000); and Alexeev et al., Nature Biotechnology 18:43-47, 2000.

The foregoing description represents only certain embodiments and technical features of the invention. It should be understood that persons of ordinary skill in the art could make various modifications and substitutions without departing from the spirit of this invention (e.g., modification of the CMV sequence to correct other mutations in the dystrophin gene; substitution of other lipids for the FuGENE™ 6 lipid; modification of the transfection method and substitution of transfection agents). In particular, all combinations of the embodiments and technical features described herein are also considered to be within the scope of the invention.

The appended claims describe what are considered patentable aspects of the invention. But although the claims are read in light of this specification, any particular embodiment or technical feature described in this specification would not limit those claims unless it was also

explicitly recited therein. Therefore, legal protection for this invention can only be determined by reference to the issued claims and equivalents thereof with the proviso that the prior art is excluded from coverage.

5 All patents, patent applications, books, and other references cited herein are indicative of the level of skill in the art and are incorporated by reference where they are cited.

We Claim:

1. A composition for the correction of a mutated dystrophin gene comprising an oligonucleobase having both ribo-type and deoxyribo-type nucleobases, which oligonucleobase
5 comprises:

a) a first and a second homologous region that are each at least eight nucleobases in length and together at least 20 and not more than 60 nucleobases in length, in which the homologous regions are, respectively, homologous to a first fragment and a second fragment of an exon of human dystrophin or of such exon and its 5' or 3' flanking intron, in which each
10 homologous region comprises at least three nucleobases of hybrid-duplex, and

b) a heterologous region that is disposed between the first and second homologous region;

wherein the composition is effective in correcting the mutated dystrophin gene in at least some muscle cells by *in vivo* administration.

15

2. The composition of claim 1 further comprising a lipid effective in introducing the oligonucleobase into at least some muscle cells by *in vivo* administration.

3. The composition of claim 2 which consists essentially of the oligonucleobase
20 and FUGENE™ 6 lipid.

4. The composition of claim 1, wherein the oligonucleobase is linked by a covalent linker to a ligand that targets the oligonucleobase to a muscle cell.

25 5. A method of correcting a mutation in the dystrophin gene of muscle tissue in an affected subject, which comprises:

providing a composition comprising an oligonucleobase having both ribo-type and deoxyribo-type nucleobases, which oligonucleobase comprises:

a) a first and a second homologous region that are each at least eight nucleobases
30 in length and together at least 20 and not more than 60 nucleobases in length, in which the homologous regions are, respectively, homologous to a first fragment and a second fragment of the dystrophin gene of the subject, which fragments

are each adjacent to the point mutation, and in which each homologous region comprises at least three nucleobases of hybrid-duplex, and

- b) a heterologous region that is disposed between the first and second homologous region; and

5 administering to the subject an amount of the composition that is effective *in vivo* to correct the mutation in at least some muscle cells of the subject.

6. The method of claim 5, wherein the composition further comprises a lipid effective in introducing the oligonucleobase into at least some muscle cells by *in vivo*
10 administration.

7. The method of claim 6, wherein the composition consists essentially of the oligonucleobase and FUGENE™ 6 lipid.

15 8. The method of claim 5, wherein the first and second fragment are fragments of an exon of the dystrophin gene or of such exon and the 3' or 5' flanking intron of the exon.

9. The method of claim 5, wherein the composition is administered to the subject by intra-muscular injection.

20 10. The method of claim 5, wherein the oligonucleobase is linked by a covalent linker to a ligand that targets the oligonucleobase to a muscle cell.

25 11. The method of claim 5, wherein the subject is canine or murine.

12. The method of claim 5, wherein the subject is a human and the mutation is corrected in somatic cells without effecting the germline.

30 13. A method of correcting an inherited or acquired mutation in affected cells of a subject, which comprises:

providing a composition comprising an oligonucleobase having both ribo-type and deoxyribo-type nucleobases, which oligonucleobase comprises:

- 5
- a) a first and a second homologous region that are each at least eight nucleobases in length and together at least 20 and not more than 60 nucleobases in length, in which the homologous regions are, respectively, homologous to a first fragment and a second fragment of a gene with the inherited or acquired mutation, and in which each homologous region comprises at least three nucleobases of hybrid-duplex, and
 - b) a heterologous region that is disposed between the first and second homologous region; and

10 administering to the subject an amount of the composition that is effective *in vivo* to correct the mutation in at least some cells of the subject's affected tissue.

15 14. The method of claim 13, wherein the composition further comprises a lipid effective in introducing the oligonucleobase into at least some muscle cells by *in vivo* administration.

15 15. The method of claim 14, wherein the composition consists essentially of the oligonucleobase and FUGENE™ 6 lipid.

20 16. The method of claim 13, wherein the first and second fragment are fragments of an exon of the dystrophin gene or of such exon and the 3' or 5' flanking intron of the exon.

17. The method of claim 13, wherein the composition is administered to the subject by intra-muscular injection.

25 18. The method of claim 13, wherein the oligonucleobase is linked by a covalent linker to a ligand that targets the oligonucleobase to a muscle cell.

19. The method of claim 13, wherein the subject is canine or murine.

30 20. The method of claim 13, wherein the subject is a human and the mutation is corrected in somatic cells without effecting the germline.

ABSTRACT OF THE DISCLOSURE

This invention relates to the field of muscular dystrophy and methods for its treatment in humans. This invention also concerns art-recognized animal models of Duchenne muscular dystrophy in dogs (GRMD) and mice (*mdx*). Another aspect concerns chimeric mutational
5 vectors capable of inducing reversion of genetic mutations (i.e., gene repair) causing genetic disease by direct injection into affected tissue. Thus, more generally, the invention envisions direct injection of chimeric mutational vectors into affected tissues to effect gene repair therein.

FIGURE 1

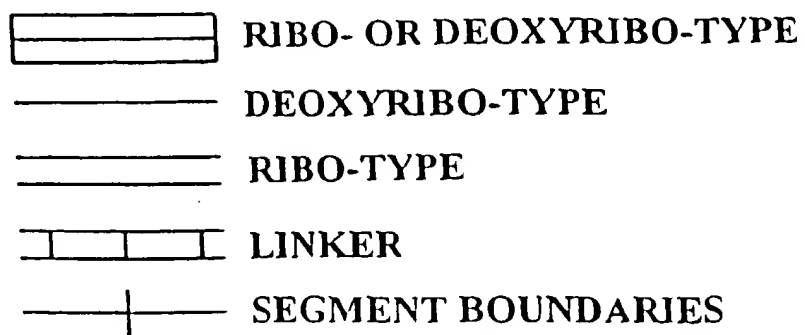
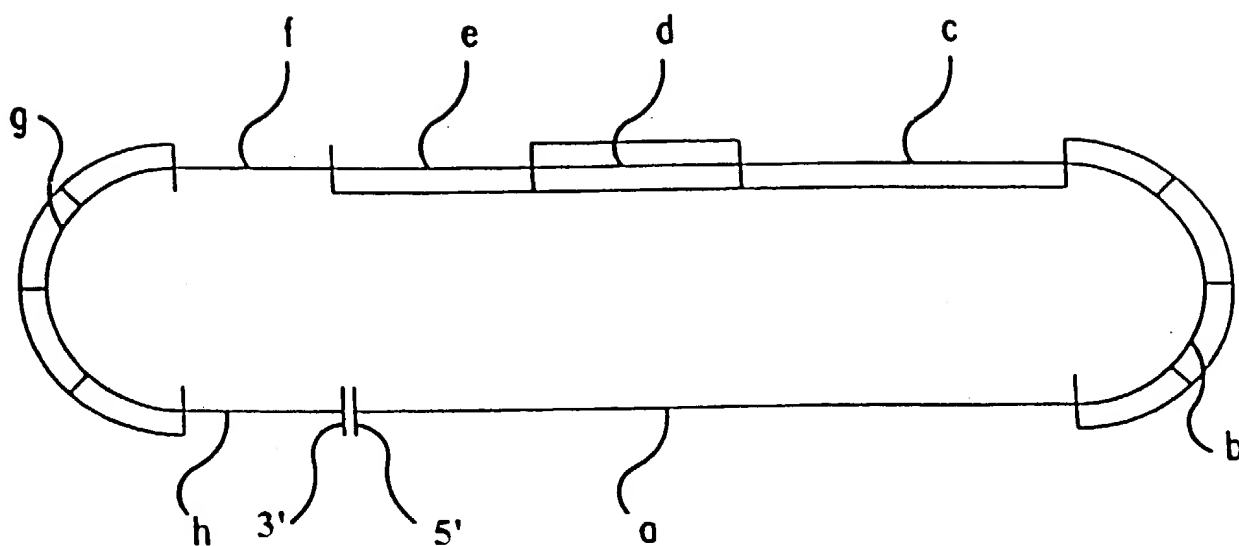


FIGURE 2

	intron 6 / EXON 7																											
Normal Human	---	t	g	t	g	t	g	t	t	t	a	g	G	C	C	A	G	A	C	C	T	C	T	T	T	G	---	
Normal Canine	---	t	g	t	g	t	g	t	t	t	c	a	g	G	C	C	A	G	A	C	C	T	C	T	T	T	G	---
GRMD Mutant	---	t	g	t	g	t	g	t	t	t	c	g	g	g	c	c	a	g	a	c	c	t	c	t	t	t	g	---

T	C	G	C	G	C	-	a	c	a	c	a	c	a	a	A	A	T	C	C	g	g	u	c	u	g	g	a	a	a	a	C	T	T
T	G	C	G	C	G		T	G	T	G	T	G	T	T	T	T	A	G	G	C	C	A	G	A	C	C	T	C	T	T	T	G	T

3'
5'

FIGURE 3

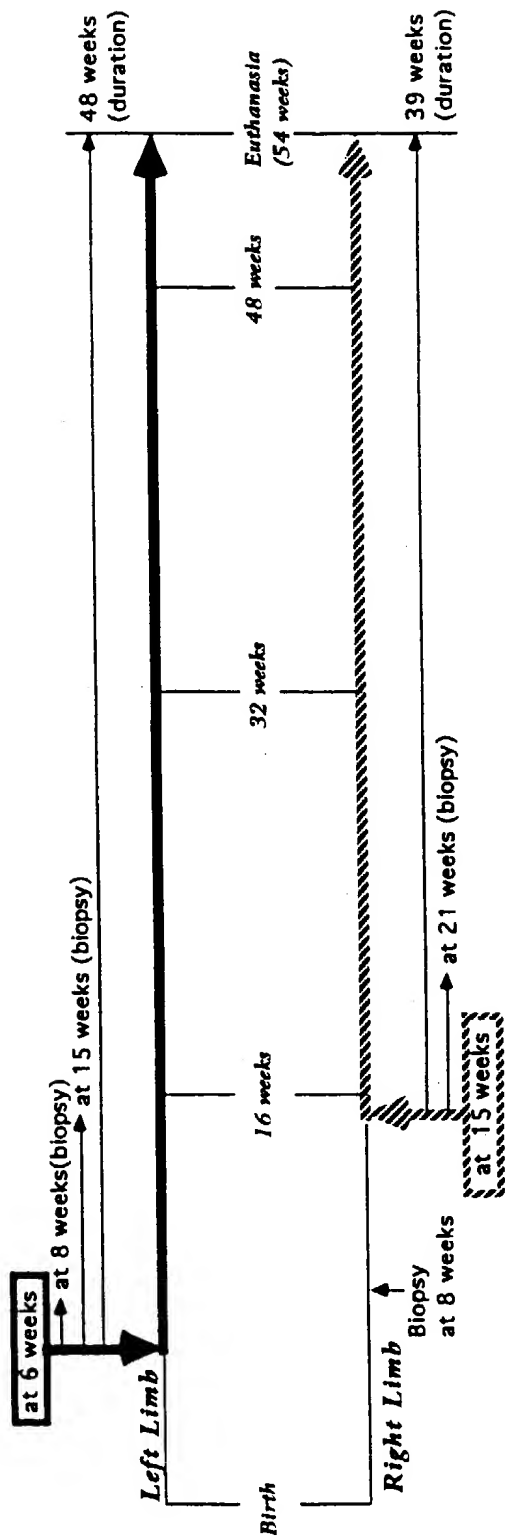


FIGURE 4

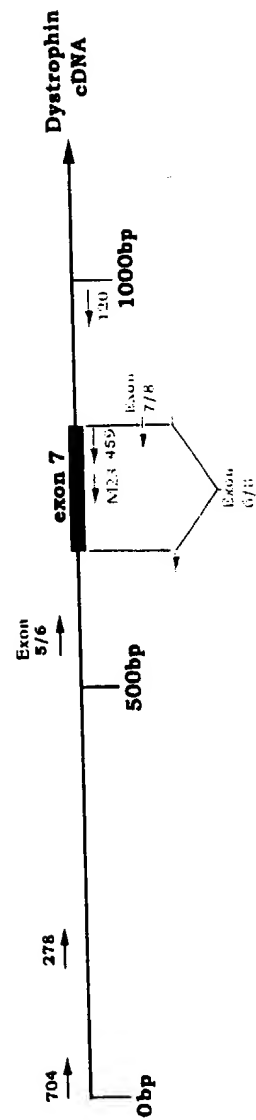


FIGURE 5

A.

normal (C57)

3170 3180 3190 3200
 ...CAA AGT TCT TTG AAA GAG CAA CAA AAT GGC TTC AAC TAT CTG AGT ...

mdx

(stop)
 ...CAA AGT TCT TTG AAA GAG CAA TAA AAT GGC TTC AAC TAT CTG AGT ...

B.

5' TCTTTGAAAGAGCAACAAAATGGCTTCAACTTTTguugaagccauuTTGTTGcucuuucaaagaCGCGCTTTTGCGCG 3'

3' 5'
 T GC GCG TCT TTG AAA GAG CAA CAA AAT GGC TTC AAC T
 T
 T
 T CG CGC aga aac uuu cuc GTT GTT uua ccg aag uug T

C.

3' 5'
 T GC GCG TCT TTG AAA GAG CAA CAA AAT GGC TTC AAC T
 T
 T
 T CG CGC aga aac uuu cuc GTT GTT uua ccg aag uug T
 ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 ...CAA AGT TCT TTG AAA GAG CAA TAA AAT GGC TTC AAC TAT CTG ...
 (mdx)



...CAA AGT TCT TTG AAA GAG CAA CAA AAT GGC TTC AAC TAT CTG ...
 ("corrected" mdx)

Pending Application 09/429,292

Filed on 10/28/1999

Inventors: Ramesh *et al.*

Our file: 7991-054-999

ATTORNEY DOCKET NUMBER: 7991-086-99

SERIAL NUMBER: 09/685,403

REFERENCE: BA

IMPROVED CHIMERIC MUTATIONAL VECTORS

This is a continuation of application Serial Number 09/078,063, filed May 12, 1998.

1. FIELD OF THE INVENTION

The invention concerns the use of duplex oligonucleobase compounds (hereafter "duplex mutational vectors") to specifically make alterations in the sequence of a DNA in a cell. In one embodiment the invention concerns compounds and methods of their use to make specific genetic alterations in the genome and in episomes (plasmids) of target prokaryotic cells. In a further embodiment the invention concerns methods of using bacterial cells to develop more efficient duplex mutational vectors. The structure of the duplex mutational vector (DMV) is designed so that genetic exchange between the DMV and the target gene occurs, i.e., a sequence contained in the DMV replaces the sequence of the target gene. In still further embodiments the invention concerns specific generic structures of DMV.

2. BACKGROUND OF THE INVENTION

U.S. Patent No. 5,565,350, issued October 15, 1996, and No. 5,731,181, issued March 24, 1998 by E.B. Kmiec, described Chimeric Mutational Vectors (CMV), i.e., vectors having both DNA-type and RNA-type nucleobases for the introduction of genetic changes in eukaryotic cells. Such CMV were characterized by having at least 3 contiguous base pairs wherein DNA-type and RNA-type nucleobases are Watson-Crick paired with each other to form a hybrid-duplex. A CMV designed to repair a mutation in the gene encoding liver/bone/kidney type alkaline phosphatase was reported in Yoon, K., et al., March 1996, Proc. Natl. Acad. Sci. **93**, 2071. The alkaline phosphatase gene was transiently introduced into CHO cells by a plasmid. Six hours later the CMV was introduced. The plasmid was recovered at 24 hours after introduction of the CMV and analyzed. The results showed that approximately 30 to 38% of the alkaline phosphatase genes were repaired by the CMV.

A CMV designed to correct the mutation in the human β -globin gene that causes Sickle Cell Disease and its successful use was described in Cole-Strauss, A., et al., 1996, Science 273:1386. A CMV designed to create a mutation in a rat blood coagulation factor IX gene in the hepatocyte of a rat is disclosed in Kren et al., 1998, Nature Medicine **4**, 285-290.

An example of a CMV having one base of a first strand that is paired with a non-complementary base of a second strand is shown in Kren et al., June 1997, *Hepatology* **25**, 1462.

United States Patent Application Serial No. 08/640,517, filed May 1, 1996, by E.B. Kmiec, A. Cole-Strauss and K. Yoon, published as WO97/41141, November 6, 1997, and application Serial No. 08/906,265, filed August 5, 1997, disclose methods and CMV that are useful in the treatment of genetic diseases of hematopoietic cells, e.g., Sickle Cell Disease, Thalassemia and Gaucher Disease.

The above-cited scientific publications of Yoon, Cole-Strauss and Kren describe CMV having two 2'-O-methyl RNA segments separated by an intervening DNA segment, which were located on the strand opposite the strand having the 5' end nucleotide. U.S. Patent No. 5,565,350 described a CMV having a single segment of 2'-O-methylated RNA, which was located on the chain having the 5' end nucleotide. An oligonucleotide having complementary deoxyribonucleotides and a continuous segment of unmodified ribonucleotides on the strand opposite the strand having the 5' end nucleotide was described in Kmiec, E.B., et al., 1994, *Mol. and Cell. Biol.* **14**:7163-7172. The sequence of the strand was derived from the bacteriophage M13mp19,

The use of single stranded oligonucleotides to introduce specific mutations in yeast are disclosed in Yamamoto, T., et al., 1992, *Genetics* **131**, 811-819. The oligonucleotides were between about 30 and 50 bases. Similar results were reported by Campbell, C.R., et al., 1989, *The New Biologist*, **1**, 223-227. Duplex DNA fragments of about 160 base pairs in length have been reported to introduce specific mutations in cultured mammalian cells. Hunger-Bertling, K., et al., 1990, *Molecular and Cellular Biochemistry* **92**, 107-116.

Applicants are aware of the following provisional applications that contain teaching with regard to uses and delivery systems of recombinagenic oligonucleotides: By Steer et al., Serial No. 60/045,288 filed April 30, 1997; Serial No. 60/054,837 filed August 5, 1997; Serial No. 60/064,996, filed November 10, 1997; and by Steer & Roy-Chowdhury et al., Serial No. 60/074,497, filed February 12, 1998, entitled "Methods of Prophylaxis and Treatment by Alteration of APO B and APO E Genes."

3. BRIEF DESCRIPTION OF THE FIGURES

Figure 1. An example of the conformation of a double hairpin type recombinagenic oligomer. The features are: a, first strand; b, second strand; c, first chain of the second strand; 1, 5' most nucleobase; 2, 3' end nucleobase; 3, 5' end nucleobase; 4, 3' most nucleobase; 5, first terminal nucleobase; 6, second terminal nucleobase.

Figure 2. An example of the conformation of a single hairpin type recombinagenic nucleobase with an overhang. The features are as above with the addition of d, the overhang. Note that the same nucleobase is both the 5' most nucleobase of the second strand and the 5' end nucleobase.

4. DEFINITIONS

The invention is to be understood in accordance with the following definitions.

An oligonucleobase is a polymer of nucleobases, which polymer can hybridize by Watson-Crick base pairing to a DNA having the complementary sequence.

Nucleobases comprise a base, which is a purine, pyrimidine, or a derivative or analog thereof. Nucleobases include peptide nucleobases, the subunits of peptide nucleic acids, and morpholine nucleobases as well as nucleobases that contain a pentosefuranosyl moiety, e.g., an optionally substituted riboside or 2'-deoxyriboside. Nucleotides are pentosefuranosyl containing nucleobases that are linked by phosphodiester. Other pentosefuranosyl containing nucleobases can be linked by substituted phosphodiester, e.g., phosphorothioate or triesterified phosphates.

A oligonucleobase compound has a single 5' and 3' end nucleobase, which are the ultimate nucleobases of the polymer. Nucleobases are either deoxyribo-type or ribo-type. Ribo-type nucleobases are pentosefuranosyl containing nucleobases wherein the 2' carbon is a methylene substituted with a hydroxyl, substituted oxygen or a halogen. Deoxyribo-type nucleobases are nucleobases other than ribo-type nucleobases and include all nucleobases that do not contain a pentosefuranosyl moiety, e.g., peptide nucleic acids.

An oligonucleobase strand generically includes regions or segments of oligonucleobase compounds that are hybridized to substantially all of the nucleobases of a complementary strand of equal length. An oligonucleobase strand has a 3' terminal nucleobase and a 5' terminal nucleobase. The 3' terminal nucleobase of a strand hybridizes to the 5'

terminal nucleobase of the complementary strand. Two nucleobases of a strand are adjacent nucleobases if they are directly covalently linked or if they hybridize to nucleobases of the complementary strand that are directly covalently linked. An oligonucleobase strand may consist of linked nucleobases, wherein each nucleobase of the strand is covalently linked to the nucleobases adjacent to it. Alternatively a strand may be divided into two chains when two adjacent nucleobases are unlinked. The 5' (or 3') terminal nucleobase of a strand can be linked at its 5'-O (or 3'-O) to a linker, which linker is further linked to a 3' (or 5') terminus of a second oligonucleobase strand, which is complementary to the first strand, whereby the two strands form a single oligonucleobase compound. The linker can be an oligonucleotide, an oligonucleobase or other compound. The 5'-O and the 3'-O of a 5' end and 3' end nucleobase of an oligonucleobase compound can be substituted with a blocking group that protects the oligonucleobase strand. However, for example, closed circular oligonucleotides do not contain 3' or 5' end nucleotides. Note that when an oligonucleobase compound contains a divided strand the 3' and 5' end nucleobases are not the terminal nucleobases of a strand.

As used herein the terms 3' and 5' have their usual meaning. The terms "3' most nucleobase", "5' most nucleobase", "first terminal nucleobase" and "second terminal nucleobase" have special definitions. The 3' most and second terminal nucleobase are the 3' terminal nucleobases, as defined above, of complementary strands of a recombinagenic oligonucleobase. Similarly, the 5' most and first terminal nucleobase are 5' terminal nucleobases of complementary strands of a recombinagenic oligonucleobase.

5. SUMMARY OF THE INVENTION

The present invention is based on the unexpected discovery that the Chimeric Mutational Vectors described in the prior art are functional in prokaryotic cells. The invention is further based on the unexpected discovery that the presence of hybrid duplex is not essential for the activity of the mutational vector. Duplex Mutational Vectors that lack three contiguous base pairs of hybrid duplex were unexpectedly found to be effective to introduce specific genetic changes in bacteria. Such vectors are termed Non-Chimeric Mutational Vectors (NCMV). NCMV can also be used in place of CMV in eukaryotic cells.

The present invention is further based on the unexpected finding that a Chimeric Mutational Vector, having a single segment of ribo-type nucleobases located on the strand

opposite the strand having the 5' end nucleobase and 3' end nucleobase is superior to the Chimeric Mutational Vectors having two segments of ribo-type nucleobases.

The invention is yet further based on the unexpected discovery of the improved efficiency of a duplex mutational vector wherein the sequence of one strand comprises the sequence of the target gene and the sequence of the second strand comprises the desired sequence, i.e., the different sequence that the user intends to introduce in place of the target sequence. Such duplex vectors are termed Heteroduplex Mutational Vectors (HDMV). An HDMV can be either Chimeric or Non-Chimeric.

In one embodiment of a HDMV, the strand that comprises the sequence of the different desired sequence is a strand having a 3' end or a 5' end. In an alternative embodiment the strand that comprises the sequence of the different, desired sequence is comprised of no ribo-type nucleobases.

The invention is yet further based on the discovery that significant improvements in the activity can be obtained by constructing the DMV so as to protect the strands of the DMV from the action of 3' exonuclease. In one embodiment 3' exonuclease protection is provided by making the DMV resistant to the action of single strand DNase.

DMV can be used to introduce specific genetic changes in target DNA sequences in prokaryotic and eukaryotic cells or episomes thereof. Such changes can be used to create new phenotypic traits not found in nature, in a subject as a therapeutic or prophylactic intervention and as an investigational tool.

6. DETAILED DESCRIPTION OF THE INVENTION

6.1. THE GENERIC STRUCTURE OF THE CHIMERIC MUTATIONAL VECTOR

The Duplex Mutational Vectors (DMV) are comprised of polymers of nucleobases, which polymers hybridize, i.e., form Watson-Crick base pairs of purines and pyrimidines, to DNA having the appropriate sequence. Each DMV is divided into a first and a second strand of at least 12 nucleobases and not more than 75 nucleobases. In a preferred embodiment the length of the strands are each between 20 and 50 nucleobases. The strands contain regions that are complementary to each other. In a preferred embodiment the two strands are complementary to each other at every nucleobase except the nucleobases wherein the target sequence and the desired sequence differ. At least two non-overlapping regions of at least 5

nucleobases are preferred.

Nucleobases contain a base, which is either a purine or a pyrimidine or analog or derivative thereof. There are two types of nucleobases. Ribo-type nucleobases are ribonucleosides having a 2'-hydroxyl, substituted 2'-hydroxyl or 2'-halo-substituted ribose. All nucleobases other than ribo-type nucleobases are deoxyribo-type nucleobases. Thus, deoxy-type nucleobases include peptide nucleobases.

In the embodiments wherein the strands are complementary to each other at every nucleobase, the sequence of the first and second strands consists of at least two regions that are homologous to the target gene and one or more regions (the "mutator regions") that differ from the target gene and introduce the genetic change into the target gene. The mutator region is directly adjacent to homologous regions in both the 3' and 5' directions. In certain embodiments of the invention, the two homologous regions are at least three nucleobases, or at least six nucleobases or at least twelve nucleobases in length. The total length of all homologous regions is preferably at least 12 nucleobases and is preferably 16 and more preferably 20 nucleobases to about 60 nucleobases in length. Yet more preferably the total length of the homology and mutator regions together is between 25 and 45 nucleobases and most preferably between 30 and 45 nucleobases or about 35 to 40 nucleobases. Each homologous region can be between 8 and 30 nucleobases and more preferably be between 8 and 15 nucleobases and most preferably be 12 nucleobases long.

One or both strands of the DMV can optionally contain ribo-type nucleobases. In a preferred embodiment a first strand of the DMV consists of ribo-type nucleobases only while the second strand consists of deoxyribo-type nucleobases. In an alternative embodiment the first strand consists of a single segment of deoxyribo-type nucleobases interposed between two segments of ribo-type nucleobases. In said alternative embodiment the interposed segment contains the mutator region or, in the case of a HDMV, the intervening region is paired with the mutator region of the alternative strand.

Preferably the mutator region consists of 20 or fewer bases, more preferably 6 or fewer bases and most preferably 3 or fewer bases. The mutator region can be of a length different than the length of the sequence that separates the regions of the target gene homology with the homologous regions of the DMV so that an insertion or deletion of the target gene results. When the DMV is used to introduce a deletion in the target gene there is no base identifiable

as within the mutator region. Rather, the mutation is effected by the juxtaposition of the two homologous regions that are separated in the target gene. For the purposes of the invention, the length of the mutator region of a DMV that introduces a deletion in the target gene is deemed to be the length of the deletion. In one embodiment the mutator region is a deletion of from 6 to 1 bases or more preferably from 3 to 1 bases. Multiple separated mutations can be introduced by a single DMV, in which case there are multiple mutator regions in the same DMV. Alternatively multiple DMV can be used simultaneously to introduce multiple genetic changes in a single gene or, alternatively to introduce genetic changes in multiple genes of the same cell. Herein the mutator region is also termed the heterologous region. When the different desired sequence is an insertion or deletion, the sequence of both strands have the sequence of the different desired sequence.

The DMV is a single oligonucleobase compound (polymer) of between 24 and 150 nucleobases. Accordingly the DMV contains a single 3' end and a single 5' end. The first and the second strands can be linked covalently by nucleobases or by non-oligonucleobase linkers. As used herein such linkers are not regarded as a part of the strands. Accordingly, a limitation, for example that a strand contain no ribo-type nucleobases does not exclude ribo-type nucleobases from a linker attached to said strand. As used herein, Chimeric, Non-Chimeric and Heteroduplex Mutational Vectors are each types of DMV and have the above properties.

In a preferred embodiment the 3' terminal nucleobase of each strand is protected from 3' exonuclease attack. Such protection can be achieved by several techniques now known to those skilled in the art or by any technique to be developed. In one embodiment protection from 3'-exonuclease attack is achieved by linking the 3' most (terminal) nucleobase of one strand with the 5' most (terminal) nucleobase of the alternative strand by a nuclease resistant covalent linker, such as polyethylene glycol, poly-1,3-propanediol or poly-1,4-butanediol. The length of various linkers suitable for connecting two hybridized nucleic acid strands is understood by those skilled in the art. A polyethylene glycol linker having from six to three ethylene units and terminal phosphoryl moieties is suitable. Durand, M. et al., 1990, *Nucleic Acid Research* **18**, 6353; Ma, M. Y-X., et al., 1993, *Nucleic Acids Res.* **21**, 2585-2589. A preferred alternative linker is bis-phosphorylpropyl-trans-4,4'-stilbenedicarboxamide. Letsinger, R.L., et alia, 1994, *J. Am. Chem. Soc.* **116**, 811-812; Letsinger, R.L. et alia, 1995, *J. Am. Chem. Soc.* **117**, 7323-7328, which are hereby incorporated by reference. Such

linkers can be inserted into the DMV using conventional solid phase synthesis. Alternatively, the strands of the DMV can be separately synthesized and then hybridized and the interstrand linkage formed using a thiophoryl-containing stilbenedicarboxamide as described in patent publication WO 97/05284, February 13, 1997, to Letsinger R.L. et alia.

In a further alternative embodiment the linker can be a single strand oligonucleobase comprised of nuclease resistant nucleobases, e.g., a 2'-O-methyl, 2'-O-allyl or 2'-F-ribonucleotides. The tetranucleotide sequences TTTT, UUUU and UUCG and the trinucleotide sequences TTT, UUU, or UCG are particularly preferred nucleotide linkers. A linker comprising a tri or tetrathymidine oligonucleotide is not comprised of nuclease resistant nucleobases and such linker does not provide protection from 3' exonuclease attack.

In an alternative embodiment, 3'-exonuclease protection can be achieved by the modification of the 3' terminal nucleobase. If the 3' terminal nucleobase of a strand is a 3' end, then a steric protecting group can be attached by esterification to the 3'-OH, the 2'-OH or to a 2' or 3' phosphate. A suitable protecting group is a 1,2-(ω -amino)-alkyldiol or alternatively a 1,2-hydroxymethyl-(ω -amino)-alkyl. Modifications that can be made include use of an alkene or branched alkane or alkene, and substitution of the ω -amino or replacement of the ω -amino with an ω -hydroxyl. Other suitable protecting groups include a 3' end methylphosphonate, Tidd, D.M., et alia, 1989, Br. J. Cancer, **60**, 343-350; and 3'-aminohexyl, Gamper H.G., et al., 1993, Nucleic Acids Res., **21**, 145-150. Alternatively, the 3' or 5' end hydroxyls can be derivatized by conjugation with a substituted phosphorus, e.g., a methylphosphonate or phosphorothioate.

In a yet further alternative embodiment the protection of the 3'-terminal nucleobase can be achieved by making the 3'-most nucleobases of the strand nuclease resistant nucleobases. Nuclease resistant nucleobases include peptide nucleic acid nucleobases and 2' substituted ribonucleotides. Suitable substituents include the substituents taught by United States Patent No. 5,731,181, and by U.S. Patent No. 5,334,711 and No. 5,658,731 to Sproat (Sproat), which are hereby incorporated by reference, and the substituents taught by patent publications EP 629 387 and EP 679 657 (collectively, the Martin Applications), which are hereby incorporated by reference. As used herein a 2' fluoro, chloro or bromo derivative of a ribonucleotide or a ribonucleotide having a substituted 2'-O as described in the Martin Applications or Sproat is termed a "2'-Substituted Ribonucleotide." Particular preferred

embodiments of 2'-Substituted Ribonucleotides are 2'-fluoro, 2'-methoxy, 2'-propyloxy, 2'-allyloxy, 2'-hydroxyethyloxy, 2'-methoxyethyloxy, 2'-fluoropropyloxy and 2'-trifluoropropyloxy substituted ribonucleotides. In more preferred embodiments of 2'-Substituted Ribonucleotides are 2'-fluoro, 2'-methoxy, 2'-methoxyethyloxy, and 2'-allyloxy substituted nucleotides.

The term "nuclease resistant ribonucleoside" encompasses including 2'-Substituted Ribonucleotides and also all 2'-hydroxyl ribonucleosides other than ribonucleotides, e.g., ribonucleotides linked by non-phosphate or by substituted phosphodiester. Nuclease resistant deoxyribonucleosides are defined analogously. In a preferred embodiment, the DMV preferably includes at least three and more preferably six nuclease resistant ribonucleosides. In one preferred embodiment the CMV contains only nuclease resistant ribonucleosides and deoxyribonucleotides. In an alternative preferred embodiment, every other ribonucleoside is nuclease resistant.

Each DMV has a single 3' end and a single 5' end. In one embodiment the ends are the terminal nucleobases of a strand. In an alternative embodiment a strand is divided into two chains that are linked covalently through the alternative strand but not directly to each other. In embodiments wherein a strand is divided into two chains the 3' and 5' ends are Watson-Crick base paired to adjacent nucleobases of the alternative strand. In such strands the 3' and 5' ends are not terminal nucleobases. A 3' end or 5' end that is not the terminal nucleobase of a strand can be optionally substituted with a steric protector from nuclease activity as described above. In yet an alternative embodiment a terminal nucleobase of a strand is attached to a nucleobase that is not paired to a corresponding nucleobase of the opposite strand and is not a part of an interstrand linker. Such embodiment has a single "hairpin" conformation with a 3' or 5' "overhang." The unpaired nucleobase and other components of the overhang are not regarded as a part of a strand. The overhang may include self-hybridized nucleobases or non-nucleobase moieties, e.g., affinity ligands or labels. In a particular preferred embodiment of DMV having a 3' overhang, the strand containing the 5' nucleobase is composed of deoxy-type nucleobases only, which are paired with ribo-type nucleobase of the opposite strand. In a yet further preferred embodiment of DMV having a 3' overhang, the sequence of the strand containing the 5' end nucleobase is the different, desired sequence and the sequence of the strand having the overhang is the sequence of the target DNA.

A particularly preferred embodiment of the invention is a DMV wherein the two strands are not fully complementary. Rather the sequence of one strand comprises the sequence of the target DNA to be modified and the sequence of the alternative strand comprises the different, desired sequence that the user intends to introduce in place of the target sequence. It follows that at the nucleobases where the target and desired sequences differ, the bases of one strand are paired with non-complementary bases in the other strand. Such DMV are termed herein Heteroduplex Mutational Vectors (HDMV). In one preferred embodiment, the desired sequence is the sequence of a chain of a divided strand. In a second preferred embodiment, the desired sequence is found on a chain or a strand that contains no ribo-type nucleobases. In a more preferred embodiment, the desired sequence is the sequence of a chain of a divided strand, which chain contains no ribo-type nucleobases.

In yet a second particularly preferred embodiment, the first strand of the CMV does not contain an intervening segment of deoxy-type nucleobases between two segments of ribo-type nucleobases. In such embodiment, the second strand is divided into a first chain and a second chain, which first chain is comprised of no ribo-type nucleobases and the portion of the first strand paired therewith contains fewer than four and preferably no deoxyribotype nucleobases. In a preferred embodiment the first chain contains the 5' end nucleobase. A yet further preferred embodiment is a Heteroduplex Mutational Vector having a single ribo-type segment according to the above, wherein the sequence of the ribo-type segment is the target DNA sequence and the sequence of the different, desired sequence is the sequence of the first chain.

6.2. INTERNUCLEOBASE LINKAGES

The linkage between the nucleobases of the strands of a DMV can be any linkage that is compatible with the hybridization of the DMV to its target sequence. Such sequences include the conventional phosphodiester linkages found in natural nucleic acids. The organic solid phase synthesis of oligonucleotides having such nucleotides is described in U.S. Patent No. Re:34,069.

Alternatively, the internucleobase linkages can be substituted phosphodiesters, e.g., phosphorothioates, substituted phosphotriesters. Alternatively, non-phosphate, phosphorus-containing linkages can be used. U.S. Patent No. 5,476,925 to Letsinger describes

phosphoramidate linkages. The 3'-phosphoramidate linkage (3'-NP(O⁻)(O)O-5') is well suited for use in DMV because it stabilizes hybridization compared to a 5'-phosphoramidate. Non-phosphate linkages between nucleobases can also be used. U.S. Patent No. 5,489,677 describes internucleobase linkages having adjacent N and O and methods of their synthesis. The linkage 3'-ON(CH₃)CH₂-5' (methylenemethylimino) is a preferred embodiment. Other linkages suitable for use in DMV are described in U.S. Patent No. 5,731,181 to Kmiec. Nucleobases that lack a pentosefuranosyl moiety and are linked by peptide bonds can also be used in the invention. Oligonucleobases containing such so-called peptide nucleic acids (PNA) are described in U.S. Patent No. 5,539,082 to Nielsen. Methods for making PNA/nucleotide chimera are described in WO 95/14706.

An complete review of the modifications at the 2' position and of the internucleobase linkage is found in Freier, S.M., & Altmann, K-H., 1997, *Nucleic Acids Research* **25**, 4429-4443.

6.3. USES OF DUPLEX MUTATIONAL VECTORS

Duplex Mutational Vectors (DMV) and particularly Non-Chimeric Mutational Vectors can be used to introduce changes into a target DNA sequence of a cell. DMV can be used according to the teaching and for the purposes that have been described by Chimeric Mutational Vectors. See, e.g., WO 97/41141 to Kmiec and Kren, B.T., et al., 1998, *Nature Medicine* **4**, 285-290.

The invention further encompasses the use of Duplex Mutational Vectors including Chimeric Mutational Vectors in prokaryotic cells that are transformation and recombination/repair competent. Mutational Vectors can be used to make specific changes in a DNA sequence of a plasmid within a bacteria, of a bacterial gene or of a bacterial artificial chromosome (BAC). Bacterial Artificial Chromosomes have been constructed based on either the bacterial F-factor origin of replication, Shizuya, H., et al., 1992, *Nature Genetics* **6**, 8794-8797; Hosoda, F., et al., 1990, *Nucleic Acids Research* **18**, 3863-3869, or on the P-1 plasmid origin of replication, Ioannou, P.A., et al., 1994, *Nature Genetics* **6**, 84-90. Heretofore the introduction of specific genetic changes in a BAC have required the construction of a plasmid containing the change followed by two recombinational events. Yang, X.W., et al., 1997, *Nature Biotechnology* **15**, 859-865; Messerle, M., et al., 1997, *Proc.*

Natl. Acad. Sci. **94**, 14759-14763. The single copy P1 based BAC pBeloBAC11, which is commercially available from Genome Systems, St. Louis Mo., is suitable for use in this embodiment of the invention.

Use of Mutational Vectors in bacteria requires that the bacteria have functional RecA and MutS genes. The RecA function can be constitutive or can be provided by a RecA gene operably linked to an inducible promoter such as the *lac* promoter, as shown in pAC184ΔTETRecA⁺. When an inducible promoter is used, RecA need be induced only for about 1 hour prior to the cells being made transformation competent and then for about one hour after electroporation. The use of an inducible RecA is preferred for certain applications where a plasmid or a bacterial artificial chromosome may be genetically destabilized by the continuous presence of RecA. Those skilled in the art will appreciate that a dominant negative RecA mutation, such as found in DH5α is unsuitable for use in the invention. Unexpectedly, activity for Mutational Vectors cannot be restored by introduction of RecA mutants that are recombinase active but lack other functions, e.g., RecAPro67.

A Mutational Vector can be introduced into the bacteria by any means that can be used to transform bacteria with plasmid DNA. In one embodiment the chimera are introduced by electroporation. The cells can be made electroporation competent by the techniques used for plasmids. The competent bacteria are then suspended in sterile nanopure water with Mutational Vectors at a concentration of between 10 ng and 10 μg per 10⁸ bacteria. Electroporation is performed in a total volume of 40 μl.

In a preferred embodiment the DMV are introduced by electroporation into the bacteria. The DMV, at about 1-2 mg/ml, are preincubated with spermidine at between 3 nM and 200 nM at room temperature in a volume of 2-4 μl prior to mixing with the bacteria to a final volume of 40 μl and electroporated. Preferably the spermidine concentration is between 5 nM and 50 nM and most preferably is about 10 nM. Without limitation as to theory, such spermidine preincubation causes the DMV to adhere to the bacteria prior to electroporation, which is believed to cause an increased rate of directed mutation. In place of spermidine, spermine or an equivalent linear polyalkylamine can be used.

Table I below shows a comparison of the rates of directed mutation in bacteria and the rates that were obtained using a cell-free extract from HuH-7 hepatocarcinoma cell line. The extract-treated DMV are then electroporated into RecA defective bacteria and the numbers of

kanamycin resistant colonies per ampicillin resistant colony calculated. The comparison shows there to be an excellent correlation between activity in the extract and activity in the bacterial system. In particular, in both systems variants **IV** and **VIb** are superior to Kany.y and in both systems Non-Chimeric Mutational Vectors having 3' exonuclease protected termini are active. The only disparity is variant **VII**, which contains solely deoxynucleotides. Variant **VII** is active in the cell-free extract but not the bacterial system.

Deoxyoligonucleotides have also been found inactive in eukaryotic cells. Without limitation as to theory, applicants believe that the activity of variant **VII** in the cell-free system is due to the reduced amount of nucleases present in the system compared to cell-containing systems. Based on these results, bacterial chimeraplasty can be used to test variant structures of recombinagenic oligonucleobases for use in eukaryotic studies.

7. EXAMPLES

7.1. MATERIALS AND METHODS

Construction of Plasmids: All DNA fragments and vectors used in cloning were isolated by gel electrophoresis and purified using the GeneClean II Kit (BIORAD101). PCR reactions were performed as follows 1-100 ng of target or genomic DNA, 5 μ L 10X buffer with Mg^{++} (Boehringer Mannheim), 0.5 μ l of 25mM dNTPs, 2.5 Units of Taq DNA Polymerase (Boehringer Mannheim), 20 pmol of each primer were mixed in a 50 μ L volume. The cycling program was: 94°C for 5 minutes, followed by 30 cycles of 94°C for 30 sec., 55 °C for 30 sec., 72 °C for 30 sec., followed by an extension at 72°C for 7 minutes. To make pWE15Kan^s, a single T→G point mutation was introduced at nucleotide position 4021 of the pWE15 vector (Stratagene) which introduced a TAG termination codon and a new *BfaI* site within the kanamycin gene. The mutant kanamycin fragment was generated from pWE15 template using the following PCR primer sets: *Set A*= Kan3910 (5'CAGGGGATCA AGATCTGAT3' (SEQ ID No. 1)-underlined bases indicate *BglII* site) and Kan4010 (5' CCCAGTCCTAGCCGAATAG 3' (SEQ ID No. 2)) *Set B*=Kan4014 (5' TCGGCTAGGACTGGGCACA 3' (SEQ ID No. 3)-underlined bases indicate *BfaI* site and bold indicates the point mutation) and Kan4594 (5'TGATAGCGGTCCGCCACA 3' (SEQ ID No. 4)-underlined bases indicate *RsrII* site.) Following digestion of product A with *BglII* and product B with *RsrII*, both products were digested with *BfaI* and ligated together. The

resultant mutant fragment was cloned into pWE15 linearized with *Bgl*III and *Rsr*II, creating pWE15Kan^s. *E. coli* strains carrying pWE15Kan^s plasmid are kanamycin sensitive.

The mutant pBR322 plasmid, pBRT^sΔ208, contains a base deletion at position 208, which results in early termination of the tetracycline gene. The deletion was created through an overlap PCR procedure as described above. The DNA products carrying the mutations were generated using primer set A {5BR22 (5' CATCGATAAGCTTTAATGC 3' (SEQ ID No. 5)) and (3BRSPH 5' CATAGTGACTGGCATGCTGTCGGA 3' (SEQ ID No. 6))} and primer set B {3BR496 (5'GCTCATGAGCCCGAAGTGGC3' (SEQ ID No. 7)) and (5BRSPH 5' TCCGACAGCATGCCAGTCACTATG 3' (SEQ ID No. 8))}. The two products were ligated together at the created *Sph*I site. The resulting fragment was digested with *Hind*III and *Bam*HI and was used to replace the analogous region in the wildtype pBR322 plasmid. The base deletion creates an *Sph*I site at position 208. The mutant pBR322 plasmid, pBRT^sm153(G), contains a stop codon in the tetracycline gene at codon 6 and was created through an overlap Polymerase Chain Reaction (PCR) procedure using fragments mixed from PCR primer set A {(5BR22 (SEQ ID No. 5) and 3BRBfa (5'CGGCATAACCTAGCCTATGCC3' (SEQ ID No. 9))} and primer set B [(3BR496 (SEQ ID No. 7)) and 5BRBfa (5'GGCTAGGTTATGCCGGTACTG3' (SEQ ID No. 10))}. The mixed products were re-amplified using primers 5BR22 and 3BR496. The resulting product was digested with *Hind*III and *Bam*HI and was used to replace the analogous region in the wildtype pBR322 plasmid. The introduction of a G at position 153 creates a stop codon and introduces a *Bfa*I digestion site. Additionally an A→G silent mutation in the tetracycline gene at position 325 was created to enable the distinction of converted from wildtype pBR322. *E. coli* strains harboring pBRT^sΔ208 and pBRT^sm153(G) plasmids are tetracycline sensitive. pET21aT^R was prepared by cloning the *Eco*RI and *Syl*I fragments into similarly digested pET21a(+) (Novagen) vector. pET21aT^R was able to confer tetracycline resistance to *E. coli* strains. pET21aT^sΔ208 and pET21aT^sm153 were prepared by replacing the *Hind*III and *Sall* region of pET21aT^R was replaced with that of pBRT^sm153(G) and pBRΔ208, respectively. *E. coli* carrying pET21aT^sm153(G) and pET21aT^sΔ208 were sensitive to tetracycline.

Construction of pAC184ΔTETRecA⁺: The tetracycline region of pACYC184 (New

England Bio Labs) vector was removed by digestion with *AvaI* and *XbaI* and replaced by an *AvaI* and *XbaI* linker {184delTet-1 (5'TCGGAGGATCCAATCTCGAGTGCAGTCAAAC 3' (SEQ ID No. 11) annealed to 184delTet-2 (5'CTAGGTTTCAGTGCAGTTCGAGATTGGATCCT3' (SEQ ID No. 12))} to make the intermediate cloning vector pAC184ΔTET. pAC184ΔTETRecA and pAC184ΔTETRecA^m were prepared by cloning RecA or RecA^m products in to the *BclI* site of pAC184ΔTET. RecA and RecA^m inserts were prepared by PCR amplification of pUCRecA and pUCRecA^m using primers 5RecALinkBCII (5'GCGTGTATCATGCACCATATGACGATTAA3' (SEQ ID No. 13)) and 3RecALinkBclI (5'GCGTGTATCAAGGAAGCGGAAGAGCGCCCA3' (SEQ ID No. 14)). The linkers define a region that contains the LacO, regulatory region (XXX) of pUC19, and the coding regions of wildtype RecA and RecA mutant (inframe deletion-removing amino acids X to X) respectively, inframe with the first five amino acids of the LacZ gene.

Construction of pAC184ΔTETRec variants: The sequence of the coding region for the RecA mutants was previously described (REF). pAC184ΔTETRec67, pAC184ΔTETRec616 and pAC184ΔTETRec659 were made by four primer PCR reactions using primers (recAxba-rec67A, rec67B, recA^{del}, recA616A, recA616B, RecA659A, RecA659B). *XbaI*/*NdeI* fragments containing the specific mutations were cloned into the *XbaI*/*NdeI* cassette of the pAC184ΔTETRec. The positive clones were isolated and the sequence was confirmed.

Construction of pAC184ΔTETmutS: The *MutS* gene was amplified from genomic DNA isolated from *E.coli DH5α* by PCR using primers MutS5' *XbaI* (5'GCGTCTAGAGATGAGTGCAATAGAAAATTT3' (SEQ ID No. 15)) and MutS3' *AseI* (5'GCGATTAATTTACACCAGACTCTTCAAGC3' (SEQ ID No. 16)). The MutS PCR product was purified using QIAquick PCR Purification Kit (Qiagen) and ligated into pGEM®-T vector (Promega) for direct TA cloning of pGEMTmutS vector. The intact wildtype MutS coding region was confirmed by sequencing. The MutS *XbaI* and *AseI* insert was ligated to the *XbaI* and *NdeI* digested pAC184ΔTetRecA expression vector, which replaces the RecA coding region with that of MutS.

Bacteria Strains and genotypes, media, and growth conditions: *E. coli* strains used in this study include RR1, MC1061, WM1100, BMS71-18, and EMSOmutS. Cells were grown in LB broth or on LB plates (10). Where appropriate cells were grown in the presence of the following antibiotics: kanamycin (50µg/mL), ampicillin, tetracycline, chloramphenicol. For transformation with plasmid or Chimera, cells were made electrocompetent essentially as described (11). Briefly, cells were grown in LB to an OD₆₀₀ of 0.5-0.7, concentrated by centrifugation (3000Xg for 10 minutes at 4°C) to 1/10th of the original volume, and washed several times (4-5) in ice-cold sterile nanopure H₂O. In the final wash, the bacteria pellet was resuspended in water (for immediate use) or 15% glycerol (for freezing at -80°C) to 1/500th of the original volume. Electrocompetent cells were either frozen immediately or were placed on ice until electroporation (up to 24 hours).

Transfection of chimera: Electrocompetent *E. coli* strains MC1061, WM100 and RR1 containing either pWE15Kan^s (for kanamycin gene targeted conversion), pET21aT^sm153(G) or pBR322T^sΔ208 (for tetracycline gene targeted conversion) were transfected with 1-2 µg of chimeras Kany.y, Tetm153 or TETΔ208, respectively, using standard electroporation conditions, 2.5 kV, 25µF, 200 Ohms. Immediately following electroporation, cells were grown for 1 hour in the presence of 1 mL of SOC (12) at 37°C with moderate shaking. We varied the time of incubation after transformation to allow sufficient time for gene targeted conversion to occur prior to antibiotic selection. Typically, following recovery in SOC medium, the entire culture was then transferred to 4 mL of LB broth containing 10µg/mL kanamycin (Sigma) for 90 min at 37°C while shaking. 1mL of this culture was then transferred to 4mL of LB broth containing 50 µg/mL kanamycin at 37 °C for 3 hr while shaking, after which an aliquot (100 µL) plated on LB agar containing 50µg/mL kanamycin and incubated overnight at 37 °C. For each bacterial strain and for each electroporation condition, kill curves were performed, as previously described.

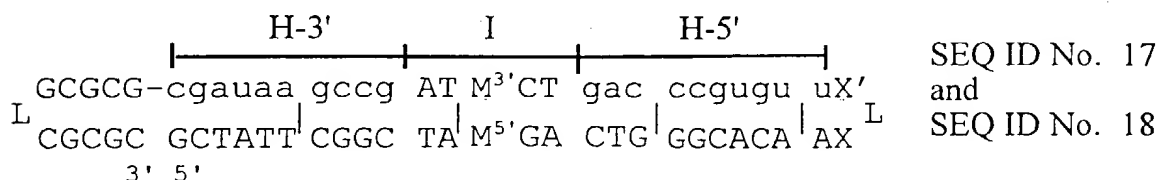
Analysis of plasmid DNA.

Plasmid DNA isolated from kanamycin resistant colonies following chimera treatment were used to transform competent DH5a bacteria. The bacteria were grown on LB plates containing ampicillin for determining total bacteria and kanamycin or tetracycline for conversion selection. Typically, from a primary isolate 3-5 secondary isolates were isolated and analyzed

by RFLP. The two populations of alleles were maintained after three replatings demonstrating that the colonies evolved from a single bacterium that contained a mixture of converted and mutant plasmids, which were subcloned and analyzed by sequence or restriction digestion.

7.2. RESULTS

The general structure of a Duplex Mutational Vector for the introduction of kanamycin resistance is given below. The intervening segment, 3' homology region, and 5' homology region are designated "I", "H-3'" and "H-5'", respectively. The interstrand linkers are designate "L". An optional chi site (5'-GCTGGTGG-3') and its complement are indicated as X and X' respectively. The 3' and 5' mutator region are single nucleotides indicated as M^{3'} and M^{5'}, respectively. Variant **I** is similar to the Chimeric Mutational Vectors described in Cole-Strauss, 1996, Science **273**, 1386, and Kren, 1998, Nature Medicine **4**, 285-290. Variant **I** is referred to as Kany.y elsewhere in this specification. The symbol "--" for a feature of a variant indicates that the feature of the variant is the same as variant **I**.



The above DMV causes a CG transversion that converts a TAG stop codon into a TAC tyr codon. Note that the first strand of **I** lacks an exonuclease protected 3' terminus and that the second strand of **I** is a divided strand, the first chain of which is the desired, different sequence. Variants **IV** and **V** are a Chimeric Mutational Vector and a Non-Chimeric Mutational Vector, respectively, having 3' termini exonuclease protected by a nuclease resistant linker (2'OMe-U₄). Variants **VIa** and **VIb** are Chimeric Heteroduplex Mutational Vectors. Variant **VIb** is the variant in which the desired, different sequence is found on the first chain, which chain consists of DNA-type nucleotides only.

The table below gives the activities of the variants relative to the variant **I** in a bacterial system and gives the frequency of conversion to kan^r/ 10⁵ plasmids for a cell-free extract. The

background rates were negligible compared to the experimental values except for variant **VIa** in the cell-free system and bacterial systems and variant **VII** in bacteria. The data reported for these variants are background corrected. Variants **VIa** and **VII** show low or absent activity. Each of variants **III-V** are superior in both systems to variant **I**, which is of the type described in the scientific publications of Yoon, Cole-Strauss and Kren cited herein above. Variant **VIII** is the optimal chimera based on inference from these data.

DMV	M ^{5'}	M ^{3'}	H5'	I	H3'	L	X(X')	Rel. Act. bacteria	kan ^r /10 ⁵ amp ^r cell-free
I	C	G	2'-OMe	DNA	2'-OMe	T ₄	None	1	6.0
II	--	--	--	--	--	--	chi†	3.2	≈1.5‡
III	--	--	--	2'-OMe	--	--	--	1.6	13
IV	--	--	--	--	--	2'-OMe-U ₄	--	10.0	50
V	--	--	DNA	--	DNA	2'-OMe-U ₄	--	3.0	9.8
VIa	G	--	--	--	--	--	--	0.06*	0.25
VIb	--	C	--	--	--	--	--	7.5	10.8
VI	--	--	--	--	--	T ₃	--	4.2	N.D.
VII	--	--	DNA	--	DNA	--	--	~0	4.4
VIII	--	C	2'-OMe	2'-OMe	2'-OMe	2'-OMe-U ₄	--	N.D.	N.D.

*Site Specific Rate

‡Result from an independent experiment normalized to other data

†GCTGGTGG

The rate of mutation can be determined by comparison of the number of kanamycin resistant (mutated) and ampicillin resistant colonies. Variant **IV** results in the mutation of a plasmid in between 1% and 2% of the viable bacteria, post electroporation, when used at between 1 µg and 2 µg of mutational vector per 10⁸ cells without the addition of spermidine on the strain MC1061. The absolute rate of mutation cannot be determined because each bacteria contains multiple copies of the pWEKan^S plasmid. For each variant, plasmid preparations were made from selected kanamycin resistant colonies, bacteria transformed and selected for kanamycin resistance. Plasmid preparations from these secondary transfectants were homogenous. Sequence of the plasmid of the secondary transfectants revealed the expected sequence in all cases except for variants **VIa** and **VII**.

The rate of conversion as a function of amount of recombinagenic oligonucleobase

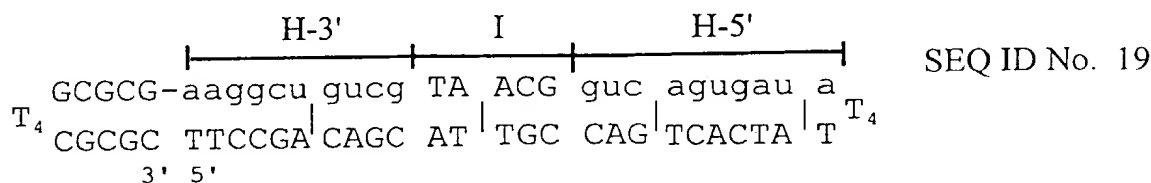
showed no maximum. Experiments using variant **I** at 0.01 $\mu\text{g}/10^8$ bacteria and 10, 100 and 1000 fold higher doses showed 5, 11, 56 and 320 converted colonies per 10^5 viable bacteria, post electroporation. The rates observed with Tet Δ 208T and Tet153 were, respectively, about 10 fold and 2 fold lower than the rate observed with variant **I** at comparable concentrations.

The preincubation of variant **I** DMV with 10 nM spermidine resulted in an approximate eight fold further increase in the number of primary kanamycin resistant colonies. An increase was also seen at 100 nM spermidine, however, no increase was apparent at 1 nM, while 1.0 mM was inhibitory.

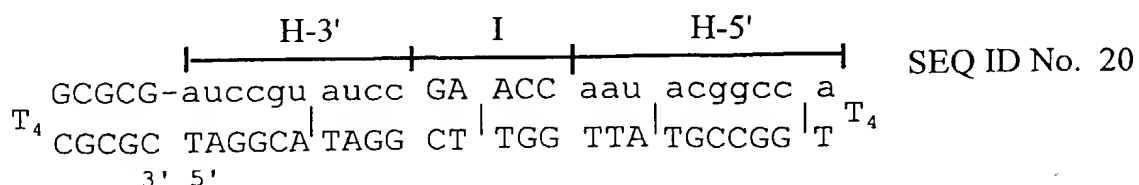
Variant **II** contains a bacterial chi site (5'GCTGGTG3') inserted between the H-5' and the linker as shown at X and X'. The replacement of the 3' most nucleotides (5'CGCGC3') by the chi site resulted in a Mutational Vector having less than a third of the activity of variant **I**.

Two tetracycline specific DMV were constructed and tested. Tet Δ 208T causes the insertion of a T that corrects a frameshift mutation. Tet 153 causes an AT transversion that converts a TAG stop codon to a TTG leu codon. The structure of tetracycline resistance Chimeric Mutational Vectors are given below

Tet Δ 208T



Tet153



SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: Kumar, Ramesh
Metz, Richard

(ii) TITLE OF THE INVENTION: Duplex Mutational Vectors
and Methods of Use Thereof In Bacterial Systems

(iii) NUMBER OF SEQUENCES: 20

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Kimeragen, Inc.
(B) STREET: 300 Pheasant Run
(C) CITY: Newtown
(D) STATE: PA
(E) COUNTRY: USA
(F) ZIP: 18940

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Hansburg, Daniel
(B) REGISTRATION NUMBER: 36156
(C) REFERENCE/DOCKET NUMBER: 7991-036-888

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: 215-504-4444
- (B) TELEFAX: 215-504-4545
- (C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CAGGGGATCA AGATCTGAT

19

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CCCAGTCCTA GCCGAATAG

19

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TCGGCTAGGA CTGGGCACA

19

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TGATAGCGGT CCGCCACA

18

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CATCGATAAG CTTTAATGC

19

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CATAGTGACT GGCATGCTGT CGGA

24

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GCTCATGAGC CCGAAGTGGC

20

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TCCGACAGCA TGCCAGTCAC TATG

24

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CGGCATAACC TAGCCTATGC C

21

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GGCTAGGTTA TGCCGGTACT G

21

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TCGGAGGATC CAATCTCGAG TGCACTGAAA C

31

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CTAGGTTTCA GTGCACTCGA GATTGGATCC T

31

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GCGTGATCAT GCACCATATG ACGATTAAA

29

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GCGTGATCAA GGAAGCGGAA GAGCGCCCA

29

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GCGTCTAGAG ATGAGTGCAA TAGAAAATTT

30

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GCGATTAATT TACACCAGAC TCTTCAAGC

29

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 84 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GCTATTCGGC TASGACTGGG CACAAGCTGG TGGTTTTCCA CCAGCTTGTG CCCAGTCSTA
GCCGAATAGC GCGCGTTTTTC GCGC

60

84

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 68 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GCTATTCGGC TASGACTGGG CACAATTTTT TGTGCCCAGT CSTAGCCGAA TAGCGCGCGT 60
TTTCGCGC 68

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 68 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TTCCGACAGC ATTGCCAGTC ACTATTTTTTA TAGTGACTGG CAATGCTGTC GGAAGCGCGT 60
TTTCGCGC 68

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 68 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TAGGCATAGG CTTGGTTATG CCGGTTTTTA CCGGCATAAC CAAGCCTATG CCTAGCGCGT 60
TTTCGCGC

WE CLAIM:

1. A heteroduplex mutational vector comprising:
 - a. a first oligonucleobase strand of at least 12 linked nucleobases and not more than 75 linked nucleobases, which strand has a first and a second terminal nucleobase;
 - b. a second oligonucleobase strand having a 3' most and a 5' most nucleobase and having a number of nucleobases equal to the first strand, which second strand is optionally divided into a first chain and a second chain; and
 - c. a single 3' end nucleobase and a single 5' end nucleobase;in which
 - i. the 3' most and 5' most nucleobases of the second strand are Watson-Crick base paired to the first terminal and the second terminal nucleobase of the first strand, respectively,
 - ii. the 3' most nucleobase of the second strand and the second terminal nucleobase of the first strand are protected from 3' exonuclease attack, and
 - iii. the second strand contains at least two non-overlapping regions of at least 5 contiguous nucleobases that are Watson-Crick base paired to nucleobases of the first strand,provided that at least one nucleobase of first strand is paired with a non-complementary base of the second strand.
2. The vector of claim 1, in which the first strand comprises not more than 50 nucleobases.
3. The vector of claim 1, in which not more than 3 nucleobases of the first strand are paired with non-complementary nucleobases of the second strand.
4. The vector of claim 3, in which the non-complementary nucleobases are deoxyribo-type nucleobases.
5. The vector of claim 3, in which not more than one nucleobase of the first strand is paired with a non-complementary nucleobase of the second strand.
6. The vector of claim 3, in which the second strand is comprised of a first chain and a second chain and the first chain contains a mismatched nucleobase.

7. The vector of claim 6, in which the first chain contains no ribo-type nucleobases.
8. The vector of claim 6, in which the first chain contains the 5' end nucleobase.
9. The vector of claim 3, in which the first strand contains at least 10 ribo-type oligonucleobases.
10. The vector of claim 9, in which the first strand contains no deoxyribo-type oligonucleobases.
11. The vector of claim 1, in which the 5' most nucleobase is linked by a nuclease resistant linker to the second terminal nucleobase, whereby said second terminal nucleobase is protected from 3' exonuclease attack.
12. The vector of claim 1, in which the 3' most nucleobase is linked by a nuclease resistant linker to the first terminal nucleobase, whereby said 3' terminal nucleobase is protected from 3' exonuclease attack.
13. The vector of claim 12, in which the 5' most nucleobase is linked by a nuclease resistant linker to the second terminal nucleobase, whereby said second terminal nucleobase is protected from 3' exonuclease attack.
14. The vector of claim 12, in which the linker comprises a moiety selected from the group consisting of 2'-methoxy-uridine, 2'-allyloxy-uridine, 2'-fluoro-uridine, 2'-methoxy-thymidine, 2'-allyloxy-thymidine, 2'-fluoro-thymidine, polyethylene glycol and trans-4,4'-stilbenecarboxamide.
15. The vector of claim 1, in which the first chain contains no ribo-type nucleobases.
16. The vector of claim 1, in which the 3' end nucleobase is protected from 3' exonuclease activity by a blocking group.
17. A chimeric duplex mutational vector having no intervening segment comprising:
 - a. a first oligonucleobase strand of at least 12 linked nucleobases and not more than 75 linked nucleobases, which strand has a first terminal and a second terminal nucleobase;
 - b. a second oligonucleobase strand having a 3' most nucleobase and a 5' most nucleobase and having a number of nucleobases equal to the first strand, which second strand is divided into a first chain and a second chain; and
 - c. a 3' end nucleobase and a 5' end nucleobase;in which

- i. the 3' most and 5' most nucleobases of the second strand are Watson-Crick base paired to the first terminal and the second terminal nucleobase of the first strand, respectively,
 - ii. the nucleobases of the first chain are deoxy-type nucleobases and nucleobases of the first strand paired therewith are nuclease resistant ribo-type nucleobases,
 - iii. the second strand contains at least two non-overlapping regions of at least 5 contiguous nucleobases that are Watson-Crick base paired to nucleobases of the first strand.
- 18. The vector of claim 17, in which the first chain comprises the 5' end nucleobase.
- 19. The vector of claim 18, in which not more than one nucleobase of the first chain is paired with a non-complementary nucleobase of the first strand.
- 20. The vector of claim 17, in which the 3' most nucleobase and the first terminal nucleobase are linked by a linker comprising a moiety selected from the group consisting of 2'-methoxy-uridine, 2'-allyloxy-uridine, 2'-fluoro-uridine, 2'-methoxy-thymidine, 2'-allyloxy-thymidine, 2'-fluoro-thymidine, polyethylene glycol and trans-4,4'-stilbenecarboxamide.
- 21. The vector of claim 17, in which the 5' most nucleobase and the second terminal nucleobase are linked by a linker comprising a moiety selected from the group consisting of 2'-methoxy-uridine, 2'-allyloxy-uridine, 2'-fluoro-uridine, 2'-methoxy-thymidine, 2'-allyloxy-thymidine, 2'-fluoro-thymidine, polyethylene glycol and trans-4,4'-stilbenecarboxamide.
- 22. An overhang containing chimeric duplex mutational vector having no intervening segment comprising:
 - a. an oligonucleobase strand of at least 12 linked nucleobases and not more than 75 linked nucleobases, which strand has a first terminal and a second terminal nucleobase; and
 - b. an oligonucleobase chain having a 3' most nucleobase and a 5' end nucleobase; and
 - c. a 3' overhang attached to the second terminal nucleobase;
 in which

- i. the 3' most and 5' end nucleobases of the chain are Watson-Crick base paired to the first terminal and the second terminal nucleobase of the first strand, respectively,
 - ii. the nucleobases of the chain are deoxy-type nucleobases and nucleobases of the strand paired therewith are nuclease resistant ribo-type nucleobases;
 - iii. the second strand contains at least two non-overlapping regions of at least 5 contiguous nucleobases that are Watson-Crick base paired to nucleobases of the first strand.
23. The vector of claim 22, in which at least one nucleobase of the first chain is paired with a non-complementary nucleobase of the first strand.
24. The vector of claim 23, in which not more than one nucleobase of the first chain is paired with a non-complementary nucleobase of the first strand.
25. The vector of claim 22, in which the 3' most nucleobase and the first terminal nucleobase are linked by a linker comprising a moiety selected from the group consisting of 2'-methoxy-uridine, 2'-allyloxy-uridine, 2'-fluoro-uridine, 2'-methoxy-thymidine, 2'-allyloxy-thymidine, 2'-fluoro-thymidine, polyethylene glycol and trans-4,4'-stilbenecarboxamide.

THE FOLLOWING CLAIMS ARE NOT SUBMITTED FOR EXAMINATION, BUT ARE INTENDED TO
MAINTAIN CONTINUITY OF DISCLOSURE ONLY FOR USE IN FUTURE APPLICATIONS.

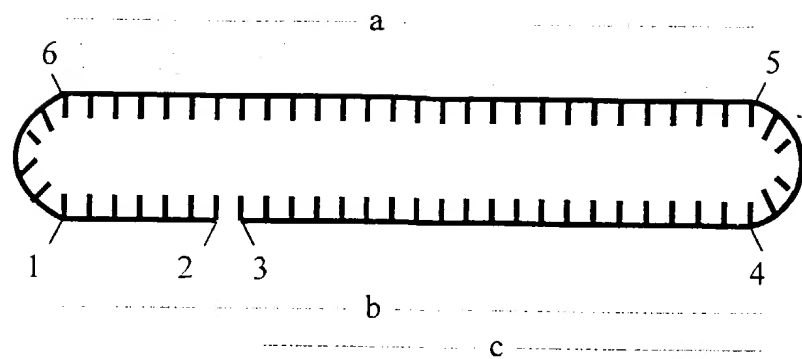
35. A method of transforming a target DNA sequence into a different, desired sequence in a bacterium that comprises introducing into the bacterium a duplex mutational vector comprising:
- a. a first oligonucleobase strand of at least 12 linked nucleobases and not more than 75 linked nucleobases, which strand has a first and a second terminal nucleobase;
 - b. a second oligonucleobase strand having an equal number of nucleobases as the first strand, which strand is optionally divided into a first chain and a second chain; and
 - c. a 3' end nucleobase and a 5' end nucleobase,
- in which
- i. the 3' most and 5' most nucleobases of the second strand are Watson-Crick base paired to the first terminal and the second terminal nucleobase, respectively, and
 - ii. the second strand contains at least two non-overlapping regions of at least 5 contiguous nucleobases that are Watson-Crick base paired to nucleobases of the first strand and the 3' end or the 5' end nucleobase;
- wherein the sequence of at least one strand comprises the different, desired sequence.
36. The method of claim 35, wherein the 3' most nucleobase of the second strand is protected from 3' exonuclease attack.
37. The method of claim 35, wherein the second terminal nucleobase of the first strand is protected from 3' exonuclease attack.
38. The method of claim 35, wherein at least one base of the first strand is paired with a non-complementary base of the second strand.
39. The method of claims 38, wherein the sequence of a strand of the mutational oligonucleobase comprises the sequence of the target DNA and the sequence of a strand of the oligonucleobase comprises the sequence of the different, desired sequence.

40. The method of claim 39, wherein the sequence of the first strand comprises the sequence of the target DNA and the second strand comprises the 5' end nucleobase.
41. The method of claim 39, wherein the sequence of the first strand comprises the sequence of the target DNA and the second strand contains no ribo-type nucleobases.
42. The method of claim 35, which further comprises the step of transiently producing functional RecA in the bacterium.
43. The method of claim 35, wherein a terminal nucleobase and a distal end nucleobase are protected from 3' exonuclease attack.
44. The method of claim 35, wherein the DNA target sequence is a sequence of a bacterial artificial chromosome or a plasmid.
45. A nucleic acid comprising RecA gene operably linked to an induceable promoter.
46. A bacterium comprising the nucleic acid of claim 45.

THE FOREGOING CLAIMS ARE NOT SUBMITTED FOR EXAMINATION, BUT ARE INTENDED TO MAINTAIN CONTINUITY OF DISCLOSURE ONLY FOR USE IN FUTURE APPLICATIONS.

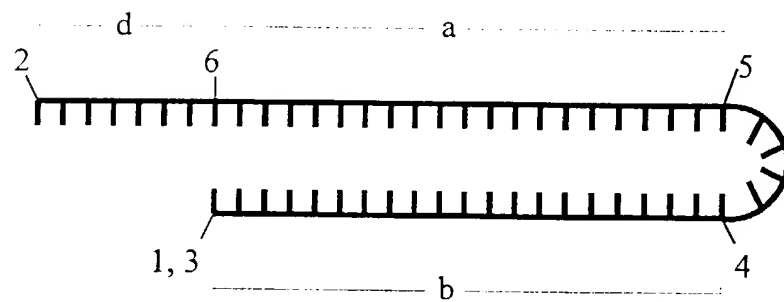
ABSTRACT

The invention is based on the discovery that recombinagenic oligonucleobases are active in prokaryotic cells that contain a strand transfer activity (*RecA*) and mismatch repair activity (*MutS*). Using this system a type of Duplex Mutational Vector termed a Heteroduplex Mutational Vector, was shown to be more active in prokaryotic cells than the types of mutational vectors heretofore tested. Further improvements in activity were obtained by replacing the tetrathymidine linker by a nuclease resistant oligonucleotide, such as tetra-2'-O-methyl-uridine, to link the two strands of the recombinagenic oligonucleobase and removing the DNA-containing intervening segment. The claims concern Duplex Mutational Vectors that contain the above improvements. In an alternative embodiment the claims concern the use of Duplex Mutational Vectors in prokaryotic cells.



Double Hairpin

Fig. 1



Single Hairpin With Overhang

Fig. 2

Pending Application 09/429,291

Filed on 10/28/1999

Inventors: Kmiec *et al.*

Our file: 7991-055-999

EUKARYOTIC USE OF IMPROVED CHIMERIC MUTATIONAL VECTORS

This is a continuation of application serial number 09/078,064, filed May 12, 1998.

1. FIELD OF THE INVENTION

Chimeraplasty concerns the introduction of directed alterations in a specific site of the DNA of a target cell by introducing duplex oligonucleotides, which are processed by the cell's homologous recombination and error repair systems so that the sequence of the target DNA is converted to that of the oligonucleotide where they are different. The present invention concerns a chimeraplasty method that is practiced in a cell-free system.

2 BACKGROUND TO THE INVENTION

2.1 CHIMERAPLASTY

Chimeraplasty in eukaryotic cells and duplex recombinagenic oligonucleotides for use therein are disclosed in U.S. Patent No. 5,565,350, issued October 15, 1996, and No. 5,731,181, issued March 24, 1998 by E.B. Kmiec (collectively "Kmiec"). The recombinagenic oligonucleotides disclosed by Kmiec contained ribo-type, e.g., 2'-O-methyl-ribonucleotides, and deoxyribo-type nucleotides that were hybridized to each other and were termed Chimeric Mutational Vectors (CMV). A CMV designed to repair a mutation in the gene encoding liver/bone/kidney type alkaline phosphatase was reported in Yoon, K., et al., 1996, Proc. Natl. Acad. Sci. **93**, 2071. The alkaline phosphatase gene was transiently introduced into CHO cells by a plasmid. Six hours later the CMV was introduced. The plasmid was recovered at 24 hours after introduction of the CMV and analyzed. The results showed that approximately 30% to 38% of the alkaline phosphatase genes were repaired by the CMV.

A CMV designed to correct the mutation in the human β -globin gene that causes Sickle Cell Disease and its successful use was described in Cole-Strauss, A., et al., 1996, Science **273**, 1386. A CMV designed to create a mutation in a rat blood coagulation factor IX gene in the hepatocyte of a rat is disclosed in Kren et al., 1998, Nature Medicine **4**, 285-290. An example of a CMV having one base of a first strand that is paired with a non-complementary base of a second strand is shown in Kren et al., June 1997, Hepatology **25**, 1462.

United States Patent Application Serial No. 08/640,517, filed May 1, 1996, by E.B. Kmiec, A. Cole-Strauss and K. Yoon, published as WO97/41141, November 6, 1997, and application Serial No. 08/906,265, filed August 5, 1997, disclose methods and CMV that are useful in the treatment of genetic diseases of hematopoietic cells, e.g., Sickle Cell Disease, Thalassemia and Gaucher Disease.

An example of the use of a CMV having one base of a first strand that is paired with a non-complementary base of a second strand is shown in Kren et al., June 1997, *Hepatology* **25**, 1462. In *Kren*, the strand having the different desired, sequence was the strand having 2'-O-methyl ribonucleotides, which was paired with the strand having the 3' end and 5' end. U.S. Patent No. 5,565,350 described a CMV having a single segment of 2'-O-methylated RNA, which was located on the chain having the 5' end nucleotide.

Applicants are aware of the following provisional applications that contain teaching with regard to chimeric mutational vectors: By Steer et al., Serial No. 60/045,288 filed April 30, 1997; Serial No. 60/054,837 filed August 5, 1997; Serial No. No. 60/064,996, filed November 10, 1997; and by Steer & Roy-Chowdhury et al., Serial No. 60/074,497, filed February 12, 1998, entitled "Methods of Prophylaxis and Treatment by Alteration of APO B and APO E Genes."

2.2 CELL-FREE RECOMBINATION

Various reports of homologous recombination using a cell-free extract have been published.

Hotta, Y., et al., 1985, *Chromosoma* **93**, 140-151 report the use of an extract of yeast, mouse spermatocytes and *Lilium* to effect homologous recombination between two mutant pBR322 plasmids. One of the plasmids was supercoiled, the second plasmid could be linearized or supercoiled. The maximum rate of recombination was less than 1%. A similar experiment using mutant defective pSV2neo and extracts of EJ cells was reported in Kucherlapati, R.S. et al., 1985, *Molecular and Cellular Biology* **5**, 714-720. The maximum rate of recombination was about 0.2%. Kucherlapati reported an absolute requirement that one of the mutant plasmids be linearized. In contrast Hotta, reported recombination between two circular plasmids, although the rate of recombination between circular and linear plasmids was higher.

The report of Jessberger, R., & Berg, P., 1991, Mol. & Cell. Biol. **11**, 445 concerns recombination catalyzed by nuclear extracts between plasmids. It stands in contrast to both of the above in two respects. The rate of recombination reported was about 20%, in contrast to rates of less than 0.5%. In addition Jessberger observed the same rate of recombination between circularized plasmids as between a circularized and a linear plasmid.

A related experiment using human nuclear extracts was reported by Lopez, B.S., et al., 1992, Nucleic Acids Research **20**, 501-506. Lopez reported recombination in a cell-free system between a linearized plasmid and an unrelated supercoiled plasmid that is not viable in the subsequent selection conditions. The linearized and supercoiled plasmid each contain a *lacZ* gene; which is a mutant in the linearized plasmid. The linearized plasmid is cut in the *lacZ* gene at a variable distance from the mutation. Homologous recombination between the site of the mutation and the cut, accordingly, results in the circularization of the plasmid that then becomes viable and the gain of *lacZ* function. Lopez reports no detectable homologous recombination when the cut and the mutation were 15 base pairs apart. Homologous recombination at a low level was observed when that distance was 27 base pairs. No further increase in the rate of homologous recombination was observed when the distance was made greater than 165 base pairs. Lopez et al., 1987, Nucleic Acids Research

2.3 *RAD51* AND *RAD52* ACTIVITY IN RECOMBINATION

Homologous recombination is the process whereby the genes of two chromosomes are exchanged. The rate of homologous recombination between two genetic loci is inversely proportional to their genetic linkage, tightly linked genes rarely recombine. In addition to its genetic function homologous recombination allows a somatic cell to repair DNA damaged by double strand breaks.

The first step in homologous recombination is believed to be synapse formation. A synapse is a DNA molecule in which one chain is hybridized to two other chains. Synapse formation requires an enzymatic activity and energy input from ATP hydrolysis. An artifactual assay in a cell-free system for the enzymatic activity believed to be required for synapse formation is "strand transfer." In a typical strand transfer assay a circular single strand DNA is combined with a linear duplex to produce a "nicked" or relaxed circular duplex and a linear single strand. The *Rad51* gene from yeast, mice and humans has been cloned and

catalyzes strand transfer. *Rad51* is believed to participate in synapse formation. Baumann, P., et al., 1996, *Cell* **87**, 757-766; Gupta, R.C., 1997, *Proc. Natl. Acad. Sci.* **94**, 463-468. The strand transfer activity is further enhanced by the presence of Rad52 protein and replication protein A. Baumann, P., & West, S.C., 1997, *EMBO J.* **16**, 5198-5206; New, J.H., et al., 1998, *Nature* **391**, 407-410; Benson, F.E., et al., 1998, *Nature* **391**, 401-404. Although RAD51 protein unlike Rec A binds to duplex DNA, Baumann & West *op cit.*; Benson, F.E., et al., *EMBO J.*, **13**, 5764-5771, in the presence of RAD52, its binding is directed toward single stranded DNA.

In yeast, *Rad51* or *Rad52* defective individuals are radiation sensitive because of an inability to repair double strand breaks. In mice, *Rad51* knock out results in embryonic lethality. Tsuzuki, T., et al., *Proc. Natl. Acad. Sci.* **93**, 6236-6240; Lin, S.D., & Hasty, P.A., *Mol. Cell. Biol.*, **16**, 7133.

2.4 CELL-FREE MISMATCH REPAIR

The intrinsic (thermodynamic) fidelity of DNA replication would lead to an unacceptably high rate of mutation without the presence of an "error correcting" mechanism. Mismatch repair is one such mechanism. In mismatch repair, duplex DNA having a base paired to a non-complementary base is processed so that one of the strands is corrected. The process involves the excision of one of the strands and its resynthesis. Reports of mismatch repair in cell-free eukaryotic systems can be found in Muster-Nassal & Kolodner, 1986, *Proc. Natl. Acad. Sci.* **83**, 7618-7622 (yeast); Glazer, P.M., et al., 1987, *Mol. Cell. Biol.* **7**, 218-224 (HeLa cell); Thomas D.C., et al., 1991, *J. Biol. Chem.*, **266**, 3744-3751 (HeLa cell); Holmes et al., 1991, *Proc. Natl. Acad. Sci.*, **87**, 5837-5841 (HeLa cell and *Drosophila*). The HeLa and *Drosophila* cell-free systems required that one strand of the mismatched duplex be nicked for full activity. By contrast, reports of repair in *Xenopus* egg extracts did not require that the mismatched duplex be nicked. Varlet, I., et al., 1990, *Proc. Natl. Acad. Sci.* **87**, 7883-7887. However, in Varlet the mismatch was repaired in a random fashion, i.e., the strands acted as templates with equal frequency.

Many of the genes required for mismatch repair in yeast and humans have been cloned based on homology with the *E. coli* mismatch repair genes. Kolodner, R., 1996, *Genes & Development* **10**, 1433-1442. Cells having defective mismatch repair genes show genetic

instability, termed Replication Error (RER), particularly evident in microsatellite DNA, and malignant transformation. Extracts of RER cells do not have mismatch repair activity. Umar, A., et al., J. Biol. Chem. 269, 14367-14370.

3. BRIEF DESCRIPTION OF THE FIGURES

Figure 1. An example of the conformation of a double hairpin type recombinagenic oligomer. The features are: a, first strand; b, second strand; c, first chain of the second strand; 1, 5' most nucleobase; 2, 3' end nucleobase; 3, 5' end nucleobase; 4, 3' most nucleobase; 5, first terminal nucleobase; 6, second terminal nucleobase.

Figure 2. An example of the conformation of a single hairpin type recombinagenic nucleobase with an overhang. The features are as above with the addition of d, the overhang. Note that the same nucleobase is both the 5' most nucleobase of the second strand and the 5' end nucleobase.

4. SUMMARY OF THE INVENTION

Chimeraplasty is an increasingly important process for the treatment of human disease and the development of useful, genetically engineered plant and animal strains. The development of improved recombinagenic oligonucleotides has been greatly facilitated by the use of bacterial testing systems, which give rapid and quantitative results as described in commonly assigned regular U.S. patent application Serial No. 09/078,063, entitled "Non-Chimeric Mutational Vectors" by R. Kumar et al., and provisional application Serial No. 60/085,191, entitled "Heteroduplex Mutational Vectors and Use Thereof in Bacteria" by Kumar et al., (hereafter collectively "*Kumar*") filed on even date herewith, which are hereby incorporated by reference in its entirety. The techniques of *Kumar* do not address whether the optimal recombinagenic oligonucleotides in bacterial systems are also optimal in eukaryotes. The prior art techniques of *in vivo* and cell-culture chimeraplasty are not designed for rapid quantitative analysis and are unable to utilize the same recombinagenic oligonucleobases and DNA targets as used in the bacterial systems.

Accordingly, an objective of the present invention is an assay that can use DNA targets and recombinagenic oligonucleobases designed for bacterial systems to rapidly evaluate the compatibility between different types of recombinagenic oligonucleotides and

the recombination and repair enzymes of different phyla, e.g., do the recombination and mismatch repair enzymes of bacteria, plants, insects and mammals have differing substrate preferences?

A further objective of the invention is an assay that can rapidly determine whether a tissue or cell line is a target for chimeraplasty, i.e, whether it contains the requisite enzymes. A yet further objective is an assay to determine what agents or treatments can alter the level of chimeraplasty activity in a cell line or tissue. A yet further objective of the invention is an assay that can determine whether a compound is an agonist or antagonist of the recombination and repair pathway. An additional objective of the invention is a practical method of making specific genetic changes in a DNA sequence in a cell-free system that is an alternative to polymerase chain reaction PCR-based methods.

The present invention meets these objectives by the unexpected discovery that chimeraplasty can be performed in a cell-free system. The components of the cell-free system are an enzyme mixture containing strand transfer activity and, optionally, a mismatch repair activity, a target DNA sequence and a recombinagenic oligonucleobase. The enzyme mixture can be made by obtaining a cell extract, or a mixture of recombinantly produced purified enzymes. The target DNA sequence is preferably a plasmid that can be used to transform an expression host such as a bacteria. In a preferred embodiment the plasmid is supercoiled. The recombinagenic oligonucleobase is any oligonucleotide or oligonucleotide derivative that can be used to introduce a site specific, predetermined genetic change in a cell. As used herein a DNA duplex consisting of more than 200 deoxyribonucleotides and no nucleotide derivatives is not a recombinagenic oligonucleobase. Typically, a recombinagenic oligonucleobase is characterized by being a duplex nucleotide, including nucleotide derivatives or non-nucleotide interstrand linkers, and having between 20 and 120 nucleobases or equivalently between 10 and 60 Watson-Crick nucleobase pairs. In a preferred embodiment, the recombinagenic oligonucleobase is substantially a duplex and contains a single 3' end and 5' end; accordingly, the strands of the duplex are covalently linked by oligonucleobase or non-oligonucleobase linkers. A further embodiment of the present invention is based on the discovery that the Non-Chimeric Mutational Vectors (NCMV), according to *Kumar*, are effective substrates for the strand transfer and repair enzymes of eukaryotic and, specifically mammalian cells. Yet further embodiments of the

invention are based on the discovery that two types of recombinagenic oligonucleobases, according to *Kumar*, Heteroduplex Mutational Vectors (HDMV) and vectors having a single segment of ribo-type nucleobases in the strand opposite the strand containing the 3' end nucleobase and 5' end nucleobase, unexpectedly give superior results when used with eukaryotic and specifically in mammalian strand transfer and repair enzymes. The term Duplex Mutational Vectors (DMV) is used herein to refer to CMV, HDMV and NCMV, collectively. Note that a HDMV can be either chimeric or non-chimeric, however, the term CMV does not encompass HDMV.

5. DETAILED DESCRIPTION OF THE INVENTION

According to the present invention a reaction is carried out in a reaction mixture containing an enzyme mixture comprising strand transfer and mismatch repair activities, a DNA target and a recombinagenic oligonucleobase. In one embodiment the DNA target is a mutated antibiotic resistance gene, e.g., *tet* or *neo* (*kan*) of a plasmid and the recombinagenic oligonucleobase is a 2'-O-methyl containing a CMV according to *Kmiec*, at about a 1:200 molar ratio. The function of the mutant *tet* or *kan* is restored by specific alteration of a single base. The reaction is terminated by phenol/chloroform extraction and the extracted plasmid electroporated into *RecA* or *MutS* defective bacteria. The extent of modification of the target DNA can be determined from the ratio of the recombinant (*kan*^r or *tet*^r) colonies to the parental type (*amp*^r). No recombinant colonies, above background, were observed when the plasmid and chimera were reacted separately and recombined after chloroform/phenol extraction. Recombinant colonies were reduced about 90% when extracts of mismatch repair deficient cells (LoVo) were used. These controls indicate that the modification, up to the point of mismatch excision is completed in the reaction mixture. The frequency of recombinant colonies was about 5 per 10⁵ parental colonies using CMV of the type described in Kren et al. Nature Medicine, 1998, 4, 285-290 and Cole-Strauss et al., 1996, Science 273, 1386 (a "Cole-Straus CMV").

As used herein a cell-free enzyme mixture is deemed to have strand transfer and mismatch repair activity when the cell-free mixture can be used to obtain the above described result.

Table I below shows the effects of multiple modifications of the Cole-Strauss CMV in

both the bacterial and cell-free eukaryotic systems. There is a very good correlation between the activity of any modification measured in each system. In particular the substitution of 2'-O-methyl uracil for thymidine in the interstrand linkers (variants **IV** and **V**), the placement of the mutator only in the 5' strand (variant **VIb**) and deletion of DNA from the 3' strand significantly improved the performance of the recombinagenic oligonucleobases in both systems.

In both systems the placement of the mutator in the 3' strand (variant **VIa**) resulted in a substantial loss of function to below one in 10^5 recombinant colonies. The frequency observed with variant **VIa** was clearly higher than background. Accordingly, as used herein a recombinagenic oligonucleobase is an oligonucleobase of the type that can provide a rate of recombination in the above cell-free system at least as high as a recombinagenic oligonucleobase made according to variant **VIa** having the same mutator sequence.

Variant **VII** with a one base mutator sequence was observed to effect recombination with a frequency of $4.4 / 10^5$. This frequency was significantly greater than that observed in the bacterial systems as well as that observed in cultured cells. Without limitation as to theory, this difference is believed to be due to the relative absence of exonucleases and endonucleases from the cell free system.

5.1 THE CELL-FREE ENZYME MIXTURE

The cell-free enzyme mixture for the practice of the invention contains the strand transfer and the mismatch repair activities. As used herein the term "cell-free enzyme mixture" indicates that the mixture excludes living cells, and preferably excludes the organelles, e.g., nuclei and mitochondria. The extent of the mismatch repair that is required in the cell-free enzyme mixture depends on the method used to detect the modification of the targeted DNA sequence and the utility.

When the modification is detected by biochemical means, e.g., restriction endonuclease digestion, the mismatch repair activity will include mismatch detection, strand cutting and excision and strand resynthesis to fill the excision and ligation. When the modification is detected in a recombination defective bacteria, e.g., *E. coli* strain DH10, the strand resynthesis and ligation activities may be omitted from the cell-free enzyme mixture. As used herein "mismatch repair activity" does not include the resynthesis and ligation

activities, which may be present in the cell-free enzyme mixture but are not required in most applications.

In certain applications, e.g., to assay the effects of modifications of the recombinagenic oligonucleobase on its efficiency with plant or mammalian enzymes, it is preferred that the mismatch repair activity be provided by the cell-free enzyme mixture. Detection by biochemical means or in a host such as a *MutS* bacteria, e.g., NR9162, which lack mismatch repair is preferred.

For certain applications, it is desirable to separate the complex of target DNA and recombinagenic oligonucleobase from the uncomplexed target DNA. Separation can be readily accomplished by introducing an affinity ligand, e.g., a biotin, onto the recombinagenic oligonucleobase. In such applications, two cell-free enzyme mixtures can be used, one before and one after the separation. The first mixture should contain only the strand transfer activity and the second need contain only the mismatch repair activity.

The cell-free enzyme mixture can be obtained as a cell extract. A procedure of Li & Kelly can be used. Li, J.J., et alia., 1985, Mol. Cell. Biol. 5, 1238-1246. The Li & Kelly procedure is a "cytoplasmic extract." The cells are mechanically disrupted in hypotonic buffer and the supernatant from centrifugation of 10 min. at 2,000xg and twice of 15 min. at 12,000xg is used. Without limitation as to theory, it is believed that the physiological cellular location of the strand transfer and mismatch repair enzymes is the nucleus but that during preparation there is sufficient loss of these enzymes from the nucleus. Crude nuclear extracts made according to Dignam et al., 1983, Nucleic Acid Research 11, 1475 are not preferred.

A cell-free enzyme mixture that lacks mismatch repair can be obtained from extracts of mutant cells having the replication error phenotype. Umar et al., 1994, J. Biol. Chem. 269, 14367. The cell line LoVo has deleted both alleles of the human *MutS* homolog (MSH2) and is suitable as a source of strand transfer activity without mismatch repair activity.

In an alternative embodiment the cell-free enzyme mixture can be a composition comprising recombinantly produced enzymes. The recombinant production of a defined enzyme allows for the addition of a known amount of the defined enzyme free of all other enzymes involved in the strand transfer and mismatch repair. When a defined enzyme is added to an extract from a cell that is deficient in that enzyme the result is a defined enzyme mixture with regard to that enzyme. The production of recombinant Rad51 can be

accomplished by the methods reported by Gupta, R.C., 1997, Proc. Natl. Acad. Sci. **94**, 463-468.

5.2 THE RECOMBINAGENIC OLIGONUCLEOBASE

Recombinagenic oligonucleobases for use in a cell-free system can be constructed according to the teaching of U.S. Patent No. No. 5,565,350 and No. 5,731,181. Additionally, recombinagenic oligonucleobases can be made according to the following.

Definitions

The invention is to be understood in accordance with the following definitions.

An oligonucleobase is a polymer of nucleobases, which polymer can hybridize by Watson-Crick base pairing to a DNA having the complementary sequence.

Nucleobases comprise a base, which is a purine, pyrimidine, or a derivative or analog thereof. Nucleobases include peptide nucleobases, the subunits of peptide nucleic acids, and morpholine nucleobases as well as nucleobases that contain a pentosefuranosyl moiety, e.g., an optionally substituted riboside or 2'-deoxyriboside. Nucleotides are pentosefuranosyl containing nucleobases that are linked by phosphodiester. Other pentosefuranosyl containing nucleobases can be linked by substituted phosphodiester, e.g., phosphorothioate or triesterified phosphates.

A oligonucleobase compound has a single 5' and 3' end nucleobase, which are the ultimate nucleobases of the polymer. Nucleobases are either deoxyribo-type or ribo-type. Ribo-type nucleobases are pentosefuranosyl containing nucleobases wherein the 2' carbon is a methylene substituted with a hydroxyl, substituted oxygen or a halogen. Deoxyribo-type nucleobases are nucleobases other than ribo-type nucleobases and include all nucleobases that do not contain a pentosefuranosyl moiety, e.g., peptide nucleic acids.

An oligonucleobase strand generically includes regions or segments of oligonucleobase compounds that are hybridized to substantially all of the nucleobases of a complementary strand of equal length. An oligonucleobase strand has a 3' most (3' terminal) nucleobase and a 5' most (5' terminal) nucleobase. The 3' most nucleobase of a strand hybridizes to the 5' most nucleobase of the complementary strand. Two nucleobases of a strand are adjacent nucleobases if they are directly covalently linked or if they hybridize to nucleobases of the complementary strand that are directly covalently linked. An

oligonucleobase strand may consist of linked nucleobases, wherein each nucleobase of the strand is covalently linked to the nucleobases adjacent to it. Alternatively a strand may be divided into two chains when two adjacent nucleobases are unlinked. The 5' (or 3') terminal nucleobase of a strand can be linked at its 5'-O (or 3'-) to a linker which linker is further linked to a 3' (or 5') terminus of a second oligonucleobase strand, which is complementary to the first strand, whereby the two strands form a single oligonucleobase compound. The linker can be an oligonucleotide, an oligonucleobase or other compound. The 5'-O and the 3'-O of a 5' end and 3' end nucleobase of an oligonucleobase compound can be substituted with a blocking group that protects the oligonucleobase strand. However, for example, closed circular oligonucleotides do not contain 3' or 5' end nucleotides. Note that when an oligonucleobase compound contains a divided strand the 3' and 5' end nucleobases are not the terminal nucleobases of a strand.

Conformation:

The Duplex Mutational Vectors (DMV) are comprised of polymers of nucleobases, which polymers hybridize, i.e., form Watson-Crick base pairs of purines and pyrimidines, to DNA having the appropriate sequence. Each DMV is divided into a first and a second strand of at least 12 nucleobases and not more than 75 nucleobases. In a preferred embodiment the length of the strands are each between 20 and 50 nucleobases. The strands contain regions that are complementary to each other. In a preferred embodiment the two strands are complementary to each other at every nucleobase except the nucleobases wherein the target sequence and the desired sequence differ. At least two non-overlapping regions of at least 5 nucleobases are preferred.

Nucleobases contain a base, which is either a purine or a pyrimidine or analog or derivative thereof. There are two types of nucleobases. Ribo-type nucleobases are ribonucleosides having a 2'-hydroxyl, substituted 2'-hydroxyl or 2'-halo-substituted ribose. All nucleobases other than ribo-type nucleobases are deoxyribo-type nucleobases. Thus, deoxy-type nucleobases include peptide nucleobases.

In the embodiments wherein the strands are complementary to each other at every nucleobase, the sequence of the first and second strands consists of at least two regions that are homologous to the target gene and one or more regions (the "mutator regions") that differ from the target gene and introduce the genetic change into the target gene. The mutator

region is directly adjacent to homologous regions in both the 3' and 5' directions. In certain embodiments of the invention, the two homologous regions are at least three nucleobases, or at least six nucleobases or at least twelve nucleobases in length. The total length of all homologous regions is preferably at least 12 nucleobases and is preferably 16 and more preferably 20 nucleobases to about 60 nucleobases in length. Yet more preferably the total length of the homology and mutator regions together is between 25 and 45 nucleobases and most preferably between 30 and 45 nucleobases or about 35 to 40 nucleobases. Each homologous region can be between 8 and 30 nucleobases and more preferably be between 8 and 15 nucleobases and most preferably be 12 nucleobases long.

One or both strands of the DMV can optionally contain ribo-type nucleobases. In a preferred embodiment a first strand of the DMV consists of ribo-type nucleobases only while the second strand consists of deoxyribo-type nucleobases. In an alternative preferred embodiment the second strand is divided into a first and second chain. The first chain contains no ribo-type nucleobases and the nucleotides of the first strand that are paired with nucleobases of first chain are ribo-type nucleobases. In an alternative embodiment the first strand consists of a single segment of deoxyribo-type nucleobases interposed between two segments of ribo-type nucleobases. In said alternative embodiment the interposed segment contains the mutator region or, in the case of a HDMV, the intervening region is paired with the mutator region of the alternative strand.

Preferably the mutator region consists of 20 or fewer bases, more preferably 6 or fewer bases and most preferably 3 or fewer bases. The mutator region can be of a length different than the length of the sequence that separates the regions of the target gene homology with the homologous regions of the DMV so that an insertion or deletion of the target gene results. When the DMV is used to introduce a deletion in the target gene there is no base identifiable as within the mutator region. Rather, the mutation is effected by the juxtaposition of the two homologous regions that are separated in the target gene. For the purposes of the invention, the length of the mutator region of a DMV that introduces a deletion in the target gene is deemed to be the length of the deletion. In one embodiment the mutator region is a deletion of from 6 to 1 bases or more preferably from 3 to 1 bases. Multiple separated mutations can be introduced by a single DMV, in which case there are multiple mutator regions in the same DMV. Alternatively multiple DMV can be used

simultaneously to introduce multiple genetic changes in a single gene or, alternatively to introduce genetic changes in multiple genes of the same cell. Herein the mutator region is also termed the heterologous region. When the different desired sequence is an insertion or deletion, the sequence of both strands have the sequence of the different desired sequence.

The DMV is a single oligonucleobase compound (polymer) of between 24 and 150 nucleobases. Accordingly the DMV contains a single 3' end and a single 5' end. The first and the second strands can be linked covalently by nucleobases or by non-oligonucleobase linkers. In a preferred embodiment the 3' terminal nucleobase of each strand is protected from 3' exonuclease attack. Such protection can be achieved by several techniques now known to those skilled in the art or by any technique to be developed.

In one embodiment protection from 3'-exonuclease attack is achieved by linking the 3' most (terminal) nucleobase of one strand with the 5' most (terminal) nucleobase of the alternative strand by a nuclease resistant covalent linker, such as polyethylene glycol, poly-1,3-propanediol or poly-1,4-butanediol. The length of various linkers suitable for connecting two hybridized nucleic acid strands is understood by those skilled in the art. A polyethylene glycol linker having from six to three ethylene units and terminal phosphoryl moieties is suitable. Durand, M. et al., 1990, *Nucleic Acid Research* **18**, 6353; Ma, M. Y-X., et al., 1993, *Nucleic Acids Res.* **21**, 2585-2589. A preferred alternative linker is bis-phosphorylpropyl-trans-4,4'-stilbenedicarboxamide. Letsinger, R.L., et alia, 1994, *J. Am. Chem. Soc.* **116**, 811-812; Letsinger, R.L. et alia, 1995, *J. Am. Chem. Soc.* **117**, 7323-7328. Such linkers can be inserted into the DMV using conventional solid phase synthesis. Alternatively, the strands of the DMV can be separately synthesized and then hybridized and the interstrand linkage formed using a thiophoryl-containing stilbenedicarboxamide as described in patent publication WO 97/05284, February 13, 1997, to Letsinger R.L. et alia.

In a further alternative embodiment the linker can be a single strand oligonucleobase comprised of nuclease resistant nucleobases, e.g., a 2'-O-methyl, 2'-O-allyl or 2'-F ribonucleotides. The tetranucleotide sequences TTTT, UUUU and UUCG and the trinucleotide sequences TTT, UUU, and UCG are particularly preferred nucleotide linkers.

In an alternative embodiment 3'-exonuclease protection can be achieved by the modification of the 3' terminal nucleobase. If the 3' terminal nucleobase of a strand is a 3' end, then a steric protecting group can be attached by esterification to the 3'-OH, the 2'-OH

or to a 2' or 3' phosphate. A suitable protecting group is a 1,2-(ω -amino)-alkyldiol or alternatively a 1,2-hydroxymethyl-(ω -amino)-alkyl. Modifications that can be made include use of an alkene or branched alkane or alkene, and substitution of the ω -amino or replacement of the ω -amino with an ω -hydroxyl. Other suitable protecting groups include a 3' end methylphosphonate, Tidd, D.M., et alia, 1989, Br. J. Cancer, **60**, 343-350; and 3'-aminohexyl, Gamper H.G., et al., 1993, Nucleic Acids Res., **21**, 145-150. Alternatively, the 3' or 5' end hydroxyls can be derivatized by conjugation with a substituted phosphorus, e.g., a methylphosphonate or phosphorothioate.

In a yet further alternative embodiment the protection of the 3'-terminal nucleobase can be achieved by making the 3'-most nucleobases of the strand nuclease resistant nucleobases. Nuclease resistant nucleobases include peptide nucleic acid nucleobases and 2' substituted ribonucleotides. Suitable substituents include the substituents taught by United States Patent No. 5,731,181, and by U.S. Patent No. 5,334,711 (Sproat), which are hereby incorporated by reference, and the substituents taught by patent publications EP 629 387 and EP 679 657 (collectively, the Martin Applications), which are hereby incorporated by reference. As used herein a 2' fluoro, chloro or bromo derivative of a ribonucleotide or a ribonucleotide having a substituted 2'-O as described in the Martin Applications or Sproat is termed a "2'-Substituted Ribonucleotide." Particular preferred embodiments of 2'-Substituted Ribonucleotides are 2'-fluoro, 2'-methoxy, 2'-propyloxy, 2'-allyloxy, 2'-hydroxyethyloxy, 2'-methoxyethyloxy, 2'-fluoropropyloxy and 2'-trifluoropropyloxy substituted ribonucleotides. In more preferred embodiments of 2'-Substituted Ribonucleotides are 2'-fluoro, 2'-methoxy, 2'-methoxyethyloxy, and 2'-allyloxy substituted nucleotides.

The term "nuclease resistant ribonucleoside" encompasses including 2'-Substituted Ribonucleotides and also all 2'-hydroxyl ribonucleosides other than ribonucleotides, e.g., ribonucleotides linked by non-phosphate or by substituted phosphodiester. Nuclease resistant deoxyribonucleosides are defined analogously. In a preferred embodiment, the DMV preferably includes at least three and more preferably six nuclease resistant ribonucleosides. In one preferred embodiment the CMV contains only nuclease resistant ribonucleosides and deoxyribonucleotides. In an alternative preferred embodiment, every other ribonucleoside is nuclease resistant.

Each DMV has a single 3' end and a single 5' end. In one embodiment the ends are

the terminal nucleobases of a strand. In an alternative embodiment, a strand is divided into two chains that are linked covalently through the alternative strand but not directly to each other. In embodiments wherein a strand is divided into two chains, the 3' and 5' ends are Watson-Crick base paired to adjacent nucleobases of the alternative strand. In such strands, the 3' and 5' ends are not terminal nucleobases. A 3' end or 5' end that is not the terminal nucleobase of a strand can be optionally substituted with a steric protector from nuclease activity as described above. In yet an alternative embodiment, a terminal nucleobase of a strand is attached to a nucleobase that is not paired to a corresponding nucleobase of the opposite strand and is not a part of an interstrand linker. Such embodiment has a single "hairpin" conformation with a 3' or 5' "overhang." The unpaired nucleobase and other components of the overhang are not regarded as a part of a strand. The overhang may include self-hybridized nucleobases or non-nucleobase moieties, e.g., affinity ligands or labels. In a particular preferred embodiment of DMV having a 3' overhang, the strand containing the 5' nucleobase is composed of deoxy-type nucleobases only, which are paired with ribo-type nucleobase of the opposite strand. In a yet further preferred embodiment of DMV having a 3' overhang, the sequence of the strand containing the 5' end nucleobase is the different, desired sequence and the sequence of the strand having the overhang is the sequence of the target DNA.

A particularly preferred embodiment of the invention is a DMV wherein the two strands are not fully complementary. Rather the sequence of one strand comprises the sequence of the target DNA to be modified and the sequence of the alternative strand comprises the different, desired sequence that the user intends to introduce in place of the target sequence. It follows that the location where the target and desired sequences differ, the bases of one strand are paired with non-complementary bases in the other strand. Such DMV are termed herein Heteroduplex Mutational Vectors (HDMV). In one preferred embodiment, the desired sequence is the sequence of a chain of a divided strand. In a second preferred embodiment, the desired sequence is found on a chain or a strand that contains no ribo-type nucleobases. In a more preferred embodiment, the desired sequence is the sequence of a chain of a divided strand, which chain contains no ribo-type nucleobases.

Internucleobase linkages

The linkage between the nucleobases of the strands of a DMV can be any linkage that

is compatible with the hybridization of the DMV to its target sequence. Such sequences include the conventional phosphodiester linkages found in natural nucleic acids. The organic solid phase synthesis of oligonucleotides having such nucleotides is described in U.S. Patent No. Re: 34,069.

Alternatively, the internucleobase linkages can be substituted phosphodiesters, e.g., phosphorothioates, substituted phosphotriesters. Alternatively, non-phosphate, phosphorus-containing linkages can be used. U.S. Patent No. 5,476,925 to Letsinger describes phosphoramidate linkages. The 3'-phosphoramidate linkage (3'-NP(O⁻)(O)O-5') is well suited for use in DMV because it stabilizes hybridization compared to a 5'-phosphoramidate. Non-phosphate linkages between nucleobases can also be used. U.S. Patent No. 5,489,677 describes internucleobase linkages having adjacent N and O and methods of their synthesis. The linkage 3'-ON(CH₃)CH₂-5' (methylenemethylimino) is a preferred embodiment. Other linkages suitable for use in DMV are described in U.S. Patent No. 5,731,181 to Kmiec. Nucleobases that lack a pentosefuranosyl moiety and are linked by peptide bonds can also be used in the invention. Oligonucleobases containing such so-called peptide nucleic acids (PNA) are described in U.S. Patent No. 5,539,082 to Nielsen. Methods for making PNA/nucleotide chimera are described in WO 95/14706.

5.3 SPECIFIC USES

Heteroduplex Mutational Vectors of the invention and Non-chimeric Mutational Vectors of the invention can be used in any eukaryotic cell in the place of the prior art Chimeric Mutational Vectors. Patent publication WO 97/41141 by Kmiec et al. teaches the use of Chimeric Mutational Vectors, *ex vivo* as do U.S. Patent No. 5,565,350 and U.S. Patent No. 5,731,181. Kren et al., 1998, Nature Medicine 4, 285 provides guidance for the use of Chimeric Mutational Vectors *in vivo*.

The recombinagenic oligonucleotides can be used in cell-free systems for several purposes, which will be apparent to those skilled in the art. Examples without limitation are as follows.

The effects of modification in the purity, chemistry, size and/or conformation of recombinagenic oligonucleotides can be rapidly and quantitatively tested in cell-free systems. The cell-free system has the further advantages that efficiency of recombination can be measured independently of the efficiency of delivery.

The cell-free system can be used to test compounds that are intended to inhibit or enhance the activity of the enzymes needed for chimeraplasty, in an alternative embodiment test for compounds that replace an enzyme of the mixture. Inhibitory compounds may be competitive or non-competitive inhibitors that act directly on the enzymes involved. Alternatively, the inhibitors can act on the cell from which an extract is made to block the synthesis or accelerate the degradation of an enzyme. These compounds may act by inducing or suppressing the synthesis of the relevant enzymes or may act by inducing post-synthetic modifications that activate or inactivate the relevant enzymes.

The cell-free system can be further used to test the relevance or particular proteins to the mechanism of chimeraplasty. Such testing can, for example without limitation be performed by use of protein-specific monoclonal antibodies to determine whether the protein in question is relevant to chimeraplasty.

A further use of the cell-free system is the specific modification of plasmid, or other isolated DNA molecules. In one embodiment of use for this purpose, the recombinogenic oligonucleobase contains an affinity ligand, such as biotin, that allows the separation of the complex with the target DNA from the uncomplexed target DNA. The chimeraplasty reaction is, in this embodiment, performed using a separate strand transfer step and a mismatch repair step. This embodiment can be used to increase the proportion of modified DNA targets, so that non-selectable modifications can be made without undue expenditure of material and effort in screening. In one embodiment, the receptor for the affinity ligand is bound to a solid phase particle so that the recombinogenic oligonucleobase/target DNA complex is attached to the particle. In the second stage of the reaction the mismatch repair activity results in the modification and release of the target DNA, whereby the supernatant of the second stage of the process is enriched for the modified plasmid.

6. EXAMPLES

Table I below shows the relative numbers of kanamycin and ampicillin resistant colonies using variants of Kany.y to correct a stop-codon causing CG transversion in the kan resistance gene.

The following materials and methods were employed to obtain these data.

Cell-Free Extracts: HuH-7 (Nakabayashi, H., et al., 1982, Cancer Res. 42, 3858) cells are

grown in DMEM supplemented with 10% fetal bovine serum to mid log phase, about 5×10^5 cells/ml. The cells are mechanically dislodged from the tissue culture flask and pelleted at 500xg. The pellet is washed in ice-cold Hypotonic Buffer with sucrose (20 mM HEPES, pH 7.5, 5 mM KCl, 1.5 mM $MgCl_2$, 1 mM DTT, 250 mM sucrose), washed in ice-cold Hypotonic Buffer without sucrose and then resuspended in Hypotonic Buffer at 6.5×10^7 cells/ml and incubated on ice for 15 min. Thereafter the cells are lysed using a Dounce homogenizer, 3-5 strokes, and thereafter incubated a further 45 min on ice. The lysate is cleared by centrifugation at 10,000xg for 10 min. and the supernatant aliquoted and stored at $-80^\circ C$ until use.

Reaction Conditions: The cell-free enzyme mixture, plasmid and DMV are reacted in a final volume of 50 μ l. The reaction buffer is 20 mM Tris, pH 7.4, 15 mM $MgCl_2$, 0.4 mM DTT, and 1.0 mM ATP. Plasmid, DMV and extract protein final concentrations are 20 μ g/ml, 20 μ g/ml and 600 μ g/ml, respectively. The reaction is run in 500 μ l "Eppendorf" tubes. The tubes are prechilled on ice and the reagents added and mixed except for the extract. The extract is then added and the reaction incubated 45 min at $37^\circ C$. The reaction is stopped by chloroform/phenol extraction. The nucleic acid is precipitated with 10% (v/v) 3M sodium acetate, pH 4.8 and 2 volumes of absolute EtOH, at $-20^\circ C$.

Bacterial Transformation: The precipitated, DMV-treated plasmid is dissolved and bacteria are transformed by electroporation according to standard techniques. After electroporation the bacteria are incubated for 1 hr in the absence of antibiotic (kanamycin) and then for 4 hours in the presence of 20% of the selective level of antibiotic.

Analysis: The effectiveness of the DMV can be ascertained from the ratio of the kanamycin resistant colonies and the ampicillin resistant colonies, which is a measure of the recovery of the plasmid and the efficiency of electroporation. The ratio given in the table below is based on data obtained after a 4 hour incubation with a sub-selective level of kanamycin. Such selective incubation results in an increase in kan^r colonies of about 100 fold. The absolute frequencies, which have been corrected for the pre-plating selection are reported.

DMV: The general structure of a Duplex Mutational Vector for the introduction of kanamycin resistance is given below. The intervening segment, 3' homology region, and 5' homology region are designated "I", "H-3" and "H-5", respectively. The interstrand linkers are designated "L". An optional chi site (5'-GCTGGTGG-3') and its complement are

indicated as X and X' respectively. The 3' and 5' mutator region are single nucleotides indicated as M^{3'} and M^{5'}, respectively. Variant **I** is similar to the Chimeric Mutational Vectors described in Cole-Strauss, 1996, Science **273**, 1386, and Kren, 1998, Nature Medicine **4**, 285-290. Variant **I** is referred to as Kany.y elsewhere in this specification. The symbol "--" for a feature of a variant indicates that the feature of the variant is the same as variant **I**.

The above DMV causes a CG transversion that converts a TAG stop codon into a TAC tyr codon. Note that the first strand of **I** lacks an exonuclease protected 3' terminus and that the second strand of **I** is a divided strand, the first chain of which is the desired, different sequence. Variants **IV** and **V** are a Chimeric Mutational Vector and a Non-Chimeric Mutational Vector, respectively, having 3' termini exonuclease protected by a nuclease resistant linker (2'OMe-U₄). Variants **VIa** and **VIb** are Chimeric Heteroduplex Mutational Vectors. Variant **VIb** is the variant in which the desired, different sequence is found on the first chain, which chain consists of DNA-type nucleotides only.

The table below gives the activities of the variants relative to the variant **I** in for a bacterial system and gives the frequency of conversion to kan^r/ 10⁵ plasmids for a cell-free extract. The background rates were negligible compared to the experimental values except for variant **VIa** in the cell-free system and bacterial systems and variant **VII** in bacteria. The data reported for these variants are background corrected. Variants **VIa** and **VII** show low or absent activity. Each of variants **III-V** are superior in both systems to variant **I**, which is of the type described in the scientific publications of Yoon, Cole-Strauss and Kren cited herein above. Variant **VIII** is the optimal chimera based on inference from these data.

The results shows an excellent correlation between activity in the cell-free extract and activity in the bacterial system. In particular, in both systems variants **IV** and **VIb** are superior to Kany.y and in both systems the Non Chimeric Mutational Vectors are active. The only disparity is variant **VII**, which contains solely deoxynucleotides. Variant **VII** is active in the cell-free extract but not the bacterial system. Deoxyoligonucleotides have also been found inactive in eukaryotic cells. Without limitation as to theory, applicants believe that the activity of variant **VII** in the cell-free system is due to the reduced amount of nucleases present in the system compared to cell-containing systems. In particular, applicants have found that a 5'-end labeled 46 nt single strand DNA was not degraded (< 1%) by the cell-free

extract in a 10 min incubation at 37°C incubation. A like result was obtained with a 46 bp 5' end labeled linear duplex DNA substrate. The reaction buffer was 2 mM ATP, 1 mM DTT, 25 mM Tris-Acetate, pH 7.15, 5 mM Mg.

TABLE I

DMV	M ^s	M ^{3'}	H5'	I	H3'	L	X(X')	R.A. (bac)	kan'/10 ⁵ amp' cell-free
I	C	G	2'-OMe	DNA	2'-OMe	T ₄	None	1	6.0
II	--	--	--	--	--	--	chi†	3.2	1.4‡.
III	--	--	--	2'-OMe	--	--	--	1.6	13
IV	--	--	--	--	--	2'-OMe-U ₄	--	10.0	50
V	--	--	DNA	--	DNA	2'-OMe-U ₄	--	3.0	9.8
VIa	G	--	--	--	--	--	--	0.06*	0.25
VIb	--	C	--	--	--	--	--	7.5	10.8
VI	--	--	--	--	--	T ₃	--	4.2	N.D.
VII	--	--	DNA	--	DNA	--	--	~0	4.4
VIII	--	C	2'-OMe	2'-OMe	2'-OMe	2'-OMe-U ₄	--	N.D.	N.D.

*Site Specific Rate

†GCTGGTGG

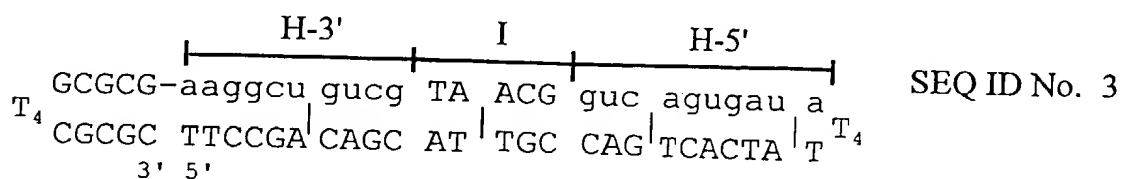
‡Result from an independent experiment normalized to other data

R.A. (bac) = relative activity (bacterial)

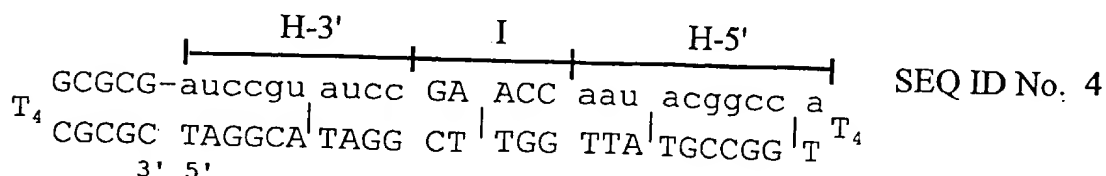
N.D. = Not Determined

The sequences of DMV for the introduction of tetracycline resistance is given below:

TetΔ208T



Tet153



SEQUENCE LISTING

(1) GENERAL INFORMATION

- (i) APPLICANT: Kmiec, Eric B.
Gamper, Howard B.
Cole-Strauss, Allyson D.
- (ii) TITLE OF THE INVENTION: EUKARYOTIC USE OF NON-CHIMERIC
MUTATIONAL VECTORS
- (iii) NUMBER OF SEQUENCES: 4
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Kimeragen, Inc.
 - (B) STREET: 300 Pheasant Run
 - (C) CITY: Newtown
 - (D) STATE: PA
 - (E) COUNTRY: USA
 - (F) ZIP: 18940
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Diskette
 - (B) COMPUTER: IBM Compatible
 - (C) OPERATING SYSTEM: DOS
 - (D) SOFTWARE: FastSEQ for Windows Version 2.0
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Hansburg, Daniel
 - (B) REGISTRATION NUMBER: 36156
 - (C) REFERENCE/DOCKET NUMBER: 7991-035-999
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 215-504-4444
 - (B) TELEFAX: 215-504-4545
 - (C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 84 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Other
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GCTATTCGGC TASGACTGGG CACAAGCTGG TGGTTTTCCA CCAGCTTGTG CCCAGTCSTA	60
GCCGAATAGC GCGCGTTTTT C GCGC	84

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 68 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GCTATTTCGGC TASGACTGGG CACAATTTTT TGTGCCCAGT CSTAGCCGAA TAGCGCGCGT	60
TTTCGCGC	68

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 68 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TTCCGACAGC ATTGCCAGTC ACTATTTTTA TAGTGACTGG CAATGCTGTC GGAAGCGCGT	60
TTTCGCGC	68

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 68 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TAGGCATAGG CTTGGTTATG CCGGTTTTTA CCGGCATAAC CAAGCCTATG CCTAGCGCGT	60
TTTCGCGC	68

WE CLAIM:

1. A method of transforming a target DNA sequence into a different, desired sequence in a eukaryotic cell that comprises:

(A) administering to the cell a chimeric duplex mutational vector comprising:

- a. a first oligonucleobase strand of at least 12 linked nucleobases and not more than 75 linked nucleobases, which strand has a first terminal and a second terminal nucleobase;
- b. a second oligonucleobase strand having a 3' most nucleobase and a 5' most nucleobase and having a number of nucleobases equal to the first strand, which second strand is divided into a first chain and a second chain; and
- c. a 3' end nucleobase and a 5' end nucleobase;

in which

- i. the 3' most and 5' most nucleobases of the second strand are Watson-Crick base paired to the first terminal and the second terminal nucleobase of the first strand, respectively,
- ii. the nucleobases of the first chain are deoxy-type nucleobases and nucleobases of the first strand paired therewith are nuclease resistant ribo-type nucleobases;
- iii. the second strand contains at least two non-overlapping regions of at least 5 contiguous nucleobases that are Watson-Crick base paired to nucleobases of the first strand,

wherein the sequence of a strand comprises the sequence of the target DNA and the sequence of a strand comprises the sequence of the different, desired sequence and the oligonucleobase segment having said different, desired sequence is comprised of at least 12 contiguous deoxyribo-type nucleobases; and

(B) detecting the presence in the cell or the progeny thereof of the DNA having the different, desired sequence.

2. The method of claim 1, wherein the vector further comprises a 3' overhang attached to the second terminal oligonucleobase, and wherein the sequence of the second strand comprises the different, desired sequence.

3. The method of claim 1, wherein the first chain comprises the 5' end nucleobase.
4. The method of claim 3, wherein not more than one nucleobase of the first chain is paired with a non-complementary nucleobase of the first strand.
5. A method of transforming a target DNA sequence into a different, desired sequence in a eukaryotic cell that comprises:
 - (A) administering to the cell a chimeric duplex mutational vector comprising:
 - a. an oligonucleobase strand of at least 12 linked nucleobases and not more than 75 linked nucleobases, which strand has a first terminal and a second terminal nucleobase; and
 - b. a oligonucleobase chain having a 3' most nucleobase and a 5' end nucleobase; and
 - c. a 3' overhang attached to the second terminal nucleobase,in which
 - i. the 3' most and 5' end nucleobases of the chain are Watson-Crick base paired to the first terminal and the second terminal nucleobase of the strand, respectively,
 - ii. the nucleobases of the chain are deoxy-type nucleobases and nucleobases of the strand paired therewith are nuclease resistant ribo-type nucleobases;
 - iii. the chain contains at least two non-overlapping regions of at least 5 contiguous nucleobases that are Watson-Crick base paired to nucleobases of the strand,wherein the sequence of the chain comprises the sequence of the different, desired sequence; and
 - (B) detecting the presence in the cell or the progeny thereof of the DNA having the different, desired sequence.
6. The method of claim 5, wherein the sequence of the strand comprises the sequence of the target DNA.
7. The method of claim 6, wherein not more than one nucleobase of the first chain is

paired with a non-complementary nucleobase of the first strand.

THE FOLLOWING CLAIMS ARE NOT SUBMITTED FOR EXAMINATION, BUT ARE INTENDED TO MAINTAIN CONTINUITY OF DISCLOSURE ONLY FOR USE IN FUTURE APPLICATIONS.

1. A cell-free composition for the modification of a DNA sequence comprising:
 - a. a duplex DNA containing a target sequence;
 - b. a recombinagenic oligonucleobase, which targets the DNA sequence and encodes the modification thereof; and
 - c. a cell-free enzyme mixture comprising a strand transfer activity.
2. The composition of claim 1, in which the oligonucleobase comprises at least 20 and not more than 200 nucleobases.
3. The composition of claim 1, in which the oligonucleobase comprises at least 10 and not more than 60 Watson-Crick nucleobase pairs.
4. The composition of claim 1, in which the oligonucleobase comprises a single 3' end and a single 5' end.
5. The composition of claim 1, in which the duplex DNA comprises two closed circular DNA polymers.
6. The composition of claim 1, in which the duplex DNA sequence is a portion of a gene-of-interest that is operably linked to a promoter, so that the gene-of-interest can be expressed in a host organism.
7. The composition of claim 6, in which the cell-free enzyme mixture lacks mismatch repair activity.
8. The composition of claim 1, in which the strand transfer activity is provided by a eukaryote-derived enzyme.

9. The composition of claim 8, in which the cell-free enzyme mixture is a defined enzyme mixture with regard to a Rad51 mammalian homolog or a Rad52 mammalian homolog.
10. The composition of claim 8, in which the cell-free enzyme mixture is an extract of a eukaryotic cell.
11. The composition of claim 10, in which the cell-free enzyme mixture is an extract of a mammalian cell.
12. The composition of claim 1, in which the cell-free enzyme mixture further comprises a mismatch repair activity.
13. The composition of claim 12, in which the mismatch repair activity is provided by a eukaryote-derived enzyme.
14. The composition of claim 13, in which the cell-free enzyme mixture is an extract of a eukaryotic cell.
15. The composition of claim 14, in which the cell-free enzyme mixture is an extract of a mammalian cell.
16. The composition of claim 12, in which the strand transfer activity is provided by a eukaryote-derived enzyme.
17. The composition of claim 16, in which the cell-free enzyme mixture is a eukaryotic cell extract.
18. The composition of claim 1, in which the recombinagenic oligonucleobase is a duplex mutational vector comprising:
 - a. a first oligonucleobase strand of at least 12 linked nucleobases and not more

- than 75 linked nucleobases, which strand has a first and a second terminal nucleobase;
- b. a second oligonucleobase strand having an equal number of nucleobases as the first strand, which strand is optionally divided into a first chain and a second chain; and
 - c. a 3' end nucleobase and a 5' end nucleobase;
- in which
- i. the 3' most and 5' most nucleobases of the second strand are Watson-Crick base paired to the first terminal and the second terminal nucleobase, respectively, and
 - ii. the second strand contains at least two non-overlapping regions of at least 5 contiguous nucleobases that are Watson-Crick base paired to nucleobases of the first strand;

wherein the sequence of at least one strand comprises the different, desired sequence.

19. A method of modifying a site of a gene-of-interest which comprises the steps of:
- a. reacting
 - i. a recombinagenic oligonucleobase, that encodes a modification of a gene-of-interest,
 - ii. a duplex DNA molecule containing the gene-of-interest operably linked to a promoter, so that the gene of interest can be expressed in a host organism, and
 - iii. a cell-free enzyme mixture comprising a strand transfer activity and a mismatch repair activity;whereby the gene-of-interest is modified at the target site;
 - b. introducing the modified gene-of-interest into the organism; and
 - c. detecting the expression of the modified gene-of-interest.
20. The method of claim 19, wherein the oligonucleobase comprises at least 20 and not more than 200 nucleobases.

21. The method of claim 19, wherein the oligonucleobase comprises at least 10 and not more than 60 Watson-Crick nucleobase pairs.
22. The method of claim 19, wherein the oligonucleobase comprises a single 3' end and a single 5' end.
23. The method of claim 19, wherein the duplex DNA comprises two closed circular DNA polymers.
24. The method of claim 19, wherein the expression of the modified gene-of-interest confers a selectable trait on the organism.
25. The method of claim 19, wherein the expression of the modified gene-of-interest confers an observable trait on the organism.
26. A method of altering a DNA sequence, which comprises the steps of:
 - a. reacting
 - i. a recombinagenic oligonucleobase, that encodes a modification of a DNA sequence,
 - ii. a duplex DNA molecule containing the sequence, and
 - iii. a cell-free enzyme mixture comprising a strand transfer activity and a mismatch repair activity;whereby the sequence is modified;
 - b. detecting the modified sequence.
27. The method of claim 26, which further comprises fractionating a cell-free composition so as to enrich the modified duplex DNA relative to the unmodified duplex DNA, prior to detecting the modified sequence.
28. The method of claim 26, wherein the oligonucleobase comprises at least 20 and not more than 200 nucleobases.

29. The method of claim 26, wherein, the oligonucleobase comprises at least 10 and not more than 60 Watson-Crick nucleobase pairs.
30. The method of claim 26, wherein the oligonucleobase comprises a single 3' end and a single 5' end.
31. The method of claim 26, in which the recombinagenic oligonucleobase is a duplex mutational vector comprising:
- a. a first oligonucleobase strand of at least 12 linked nucleobases and not more than 75 linked nucleobases, which strand has a first and a second terminal nucleobase;
 - b. a second oligonucleobase strand having an equal number of nucleobases as the first strand, which strand is optionally divided into a first chain and a second chain; and
 - c. a 3' end nucleobase and a 5' end nucleobase;
- in which
- i. the 3' most and 5' most nucleobases of the second strand are Watson-Crick base paired to the first terminal and the second terminal nucleobase, respectively, and
 - ii. the second strand contains at least two non-overlapping regions of at least 5 contiguous nucleobases that are Watson-Crick base paired to nucleobases of the first strand;
- wherein the sequence of at least one strand comprises the different, desired sequence.
42. A method of transforming a target DNA sequence into a different, desired sequence in a eukaryotic cell that comprises (A) administering to the cell a duplex mutational vector comprising:
- d. a first oligonucleobase strand of at least 12 linked nucleobases and not more than 75 linked nucleobases, which strand has a first and a second terminal nucleobase;

- e. a second oligonucleobase strand having a 3' most and a 5' most nucleobase and having a number of nucleobases equal to the first strand, which second strand is optionally divided into a first chain and a second chain; and
- f. a single 3' end nucleobase and a single 5' end nucleobase;

in which

- i. the 3' most and 5' most nucleobases of the second strand are Watson-Crick base paired to the first terminal and the second terminal nucleobase of the first strand, respectively, and
- ii. the second strand contains at least two non-overlapping regions of at least 5 contiguous nucleobases that are Watson-Crick base paired to nucleobases of the first strand;

wherein the sequence of the first strand comprises the sequence of the target DNA and the sequence of the second strand comprises the different, desired sequence; and
(B) detecting the presence in the cell or the progeny thereof of the DNA having the different, desired sequence.

- 43. The method of claim 42, wherein the first strand comprises at least 12 ribo-type nucleobases.
- 44. The method of claim 42, wherein the second strand is divided into a first chain and a second chain.
- 45. The method of claim 44, wherein the sequence of the first chain is the different, desired sequence and the first chain contains no ribo-type nucleobases.
- 46. The method of claim 45, wherein the first chain comprises the 5' end nucleobase.
- 47. The method of claim 45, wherein the first chain comprises the 3' end nucleobase.
- 48. A method of transforming a target DNA sequence into a different, desired sequence in a eukaryotic cell that comprises:

(A) administering to the cell a duplex mutational vector comprising:

- g. a first oligonucleobase strand of at least 12 linked nucleobases and not more than 75 linked nucleobases, which strand has a first and a second terminal nucleobase;
- h. a second oligonucleobase strand having a 3' most and a 5' most nucleobase and having a number of nucleobases equal to the first strand, which second strand is optionally divided into a first chain and a second chain; and
- i. a single 3' end nucleobase and a single 5' end nucleobase,

in which

- i. the 3' most and 5' most nucleobases of the second strand are Watson-Crick base paired to the first terminal and the second terminal nucleobase of the first strand, respectively, and
- ii. the second strand contains at least two non-overlapping regions of at least 5 contiguous nucleobases that are Watson-Crick base paired to nucleobases of the first strand;

wherein the sequence of a strand comprises the sequence of the target DNA and the sequence of a strand comprises the sequence of the different, desired sequence and the oligonucleobase segment having said different, desired sequence is comprised of at least 12 contiguous deoxyribo-type nucleobases; and

(B) detecting the presence in the cell or the progeny thereof of the DNA having the different, desired sequence.

- 49. The method of claim 48, wherein the sequence of the first strand comprises the sequence of the target DNA.
- 50. The method of claim 48, wherein the sequence of the first strand comprises the sequence of the different, desired sequence.
- 51. The method of claim 48, wherein the second strand is comprised of a first chain and a second chain and the first chain contains no ribo-type nucleobases.

52. The method of claim 51, wherein the sequence of the target DNA is the sequence of the first chain.
53. The method of claim 51, wherein the sequence of the different, desired sequence is the sequence of the first chain.

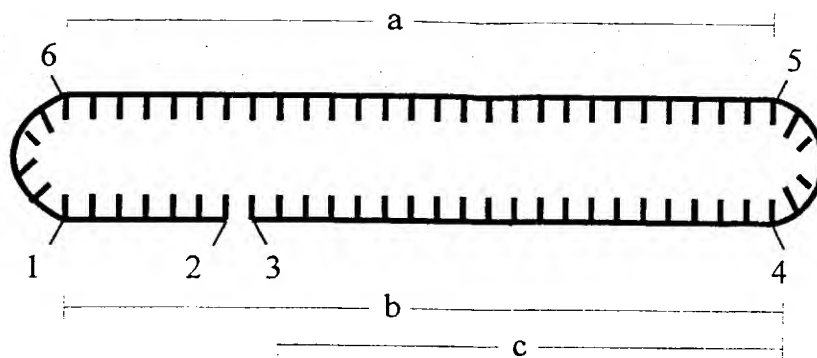
THE FORGOING CLAIMS ARE NOT SUBMITTED FOR EXAMINATION, BUT ARE INTENDED TO MAINTAIN CONTINUITY OF DISCLOSURE ONLY FOR USE IN FUTURE APPLICATIONS.

ABSTRACT

The invention is based on the reaction of recombinagenic oligonucleotides in a cell-free system containing a cytoplasmic cell extract and a test duplex DNA on a plasmid. The reaction specifically converts a mutant *kan^r* gene to recover the resistant phenotype in transformed *MutS*, *RecA* deficient bacteria and allows for the rapid and quantitative comparison of recombinagenic oligonucleobases. Using this system a type of Duplex Mutational Vector termed a Heteroduplex Mutational Vector, was shown to be more active in than the types of mutational vectors heretofore tested. Further improvements in activity were obtained by replacement of a tetrathymidine linker by a nuclease resistant oligonucleotide, such as tetra-2'-O-methyl-uridine, to link the two strands of the Duplex Mutational Vector and removal of the DNA-containing intervening segment. The claims concern Duplex Mutational Vectors that contain the above improvements. In an alternative embodiment the claims concern a reaction mixture containing a recombinagenic oligonucleobase, a cell-free enzyme mixture and a duplex DNA containing a target sequence. In yet an alternative embodiment, the invention concerns the use of such mixture to test improvements in recombinagenic oligonucleobases, as well as to test the effects of compounds on the activity of the cell-free enzyme mixture and also to make specific changes in the target DNA sequence.

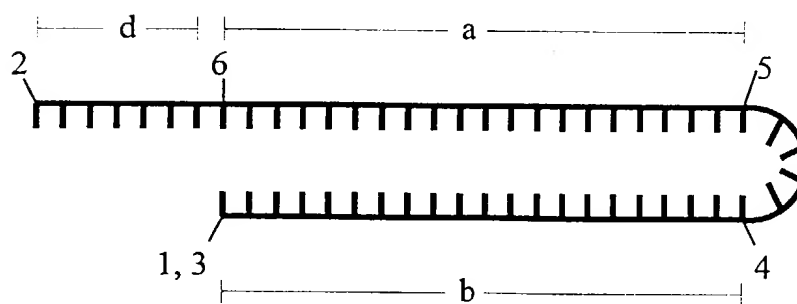
ABSTRACT

The invention is based on the reaction of recombinagenic oligonucleotides in a cell-free system containing a cytoplasmic cell extract and a test duplex DNA on a plasmid. The reaction specifically converts a mutant *kan^r* gene to recover the resistant phenotype in transformed *MutS*, *RecA* deficient bacteria and allows for the rapid and quantitative comparison of recombinagenic oligonucleobases. Using this system a type of Duplex Mutational Vector termed a Heteroduplex Mutational Vector, was shown to be more active in than the types of mutational vectors heretofore tested. Further improvements in activity were obtained by replacement of a tetrathymidine linker by a nuclease resistant oligonucleotide, such as tetra-2'-O-methyl-uridine, to link the two strands of the Duplex Mutational Vector and removal of the DNA-containing intervening segment. The claims concern Duplex Mutational Vectors that contain the above improvements. In an alternative embodiment the claims concern a reaction mixture containing a recombinagenic oligonucleobase, a cell-free enzyme mixture and a duplex DNA containing a target sequence. In yet an alternative embodiment, the invention concerns the use of such mixture to test improvements in recombinagenic oligonucleobases, as well as to test the effects of compounds on the activity of the cell-free enzyme mixture and also to make specific changes in the target DNA sequence.



Double Hairpin

Fig. 1



Single Hairpin With Overhang

Fig. 2

Pending Application 09/384,960

Filed on 08/27,1999

Inventors: Metz *et al.*

Our file: 7991-046-999

SINGLE-STRANDED OLIGODEOXYNUCLEOTIDE MUTATIONAL VECTORS

1. FIELD OF THE INVENTION

The invention concerns single-stranded oligodeoxynucleotides, and certain derivatives thereof and methods of their use for introducing a predetermined change at a predetermined location in a target gene in a living cell. The cell can be a mammalian or avian cell either in an artificial culture medium or in an organism, a bacterial cell or a plant cell.

2. BACKGROUND OF THE INVENTION

Techniques of making a predetermined change at a predetermined location in a target nucleic acid sequence of a cell have been described. These techniques utilize the cell's enzymes that concern DNA repair and homologous recombination. In these techniques an oligonucleotide or oligonucleotide analog is synthesized that contains two regions that have the sequence of the target gene that flank a region, termed a "mutator region," that differs from the target gene. In this application such oligonucleotides and analogs will be generically termed "mutational vectors." Such mutational vectors can introduce predetermined genetic changes into a target gene by a mechanism that is believed to involve homologous recombination and/or nucleotide excision and repair.

United States patents No. 5,565,350 and No. 5,731,181 to Kmiec describe mutational vectors that contain complementary strands wherein a first strand comprises ribonucleotide analogs that form Watson-Crick base pairs with deoxyribonucleotides of a second strand. Commonly assigned United States patent application Serial No. 09/078,063, filed May 12, 1998, describes certain improvements in duplex mutational vectors, including a variant in which the mutator region is present on only one of the two strands. The use of Kmiec type mutational vectors in mammalian systems is described in U.S. patent No. 5,760,012 and in conjunction with macromolecular carriers in patent publication WO 98/49350 to Kren et al., and in related United States patent application Serial No. 09/108,006. Additional descriptions of the use of Kmiec type mutational vectors can be found in scientific publications Cole-Strauss et al., 1996, Science 273, 1386, Scientific

publications concerning Kmiec type mutational vectors and macromolecular carriers include Kren et al., 1998, Nature Med. 4, 285; Bandyopadhyay et al., April 1999, J. Biol. Chem. 274, 10163.

The use of Kmiec type mutation vectors in plant cells is described in patent publications WO 99/25853 to Pioneer Hi-Bred International. WO 99/07865 to Kimeragen and WO 98/54330 to Zeneca Ltd.. Scientific publications that describe the use of Kmiec type vectors in plants are Beetham et al., July 1999 PNAS 96, 8774 and Zhu, et al., July 1999, PNAS 96, 8768.

The use of Kmiec type mutational vectors and variants thereof, which are double stranded is described in United States patent application Serial No. 09/078,063, filed May 12, 1998 to R. Kumar and R. Metz. The application of Kumar and Metz inter alia teaches that Kmiec type vectors and the variants thereof can be used in bacterial cells.

The use of single stranded oligodeoxynucleotides as mutational vectors to effect changes in a chromosomal gene in the yeast, *S. cerevisiae*, was described in reports from the laboratory of Dr. F. Sherman, Yale University. Moerschell, R.P., et al., 1988, Proc. Natl. Acad. Sci., 85, 524-528 and Yamamoto, T., et al., 1992, Yeast 8, 935-948. The optimum length of the mutational vectors used in these studies was 50 nucleotides.

An isolated report of the use of a 160 NT single and double stranded polynucleotide to attempt to make alterations in a chromosomal gene can be found at Hunger-Bertling, 1990, Mol. Cell. Bioch. 92, 107-116. The results for single stranded polynucleotides were ambiguous because only the product of the experiments using double-stranded polynucleotides were analyzed.

The use of single stranded DNA fragment of 488 bp to make specific genetic changes in the cystic fibrosis transmembrane conductance regulator gene has been

reported by Gruenet and colleagues. Goncz et al., Nov 1998, Hum. Mol. Genetics 7, 1913; Kunzelmann et al., 1996, Gene Ther. 3, 859-67.

Single stranded oligodeoxynucleotides of about 40 nucleotides in length in mammalian cells were used as a control for studies of episomal genes in which the oligodeoxynucleotide was covalently linked to a triplex forming oligonucleotide and that the oligodeoxynucleotide alone resulted in rates of predetermined genetic change of the episomal gene of about 1 per 5×10^4 , or fewer. Chan et al., April 1999, J. Biol. Chem. 74, 11541-11548, An earlier report of the use of single-stranded oligodeoxynucleotide to make predetermined changes in an episomal gene in a mammalian cell can be found in Campbell, C.R., et al., 1989, The New Biologist 1, 223-227.

One aspect of the invention concerns oligodeoxynucleotides that have been modified by the attachment of an Indocarbocyanine dye. Indocarbocyanine dyes are known as excellent fluorophores. The synthesis of blocked indocarbocyanine β cyanoethyl N,N-diisopropyl phosphoramidites that are suitable for use in solid phase nucleotide synthesis is described in United States patent No. 5,556,959 and No. 5,808,044.

A second aspect of the invention concerns a composition comprising a single stranded oligonucleotide encoding a predetermined genetic change and a macromolecular carrier that comprises a ligand for a receptor on the surface of the target cell. A composition comprising a poly-L-lysine, a ligand for the asialoglycoprotein receptor and an antisense oligodeoxynucleotide of between 21 and 24 nucleotides is described in patent publication WO 93/04701 to Wu, G.Y.

A third aspect of the invention concerns a modification of a oligodeoxynucleotide by the attachment of a 3'-3' linked nucleotide. United States patent No. 5,750,669, assigned to Hoechst A.G., teaches such a modified oligodeoxynucleotide.

3. SUMMARY OF THE INVENTION

The present invention is based on the unexpected discovery that single-stranded oligodeoxynucleotides, particularly when appropriately modified or placed in a composition with a suitable macromolecular carrier, can be as or more effective in making predetermined genetic changes to target genes in cells as the prior art, i.e., Kmiec type mutational vectors. A single stranded oligodeoxynucleotide suitable for use according to the present invention is termed hereafter a Single-Stranded Oligodeoxynucleotide Mutational Vector or a SSOMV.

In one embodiment the invention provides for a composition for use in making changes to the chromosomal genes of mammalian cells consisting of the oligodeoxynucleotide encoding the genetic change and a macromolecular carrier. The carrier can be either a polycation, an aqueous-cored lipid vesicle or a lipid nanosphere. In a further embodiment that is suitable for *in vivo* use, the carrier further comprises a ligand that binds to a cell-surface receptor that is internalized such as the asialoglycoprotein receptor, the folic acid receptor or the transferrin receptor. In preferred embodiments the oligodeoxynucleotide is modified by the attachment of 3' and 5' blocking substituents such as a 3'-3' linked cytosine nucleotide and a 5' linked indocarbocyanine dye. In an alternative embodiment the modification can consist of the replacement of the 3' most and/or 5' most internucleotide phosphodiester linkage with a non-hydrolyzeable linkage such as a phosphorothioatediester linkage or a phosphoramidate linkage.

In a second embodiment the invention provides for the modification of the 3' and 5' end nucleotides of the oligodeoxynucleotide that encodes the predetermined genetic change. The invention is further based on the unexpected discovery that certain such modifications do not block the effectiveness of the oligodeoxynucleotide to produce genetic changes. One such embodiment is the combination of a 3'-3' linked cytosine nucleotide and a 5' linked indocarbocyanine dye. So modified, the oligodeoxynucleotides are more than 50 fold more effective than unmodified oligodeoxynucleotides when used to make genetic changes in bacterial cells.

In a third embodiment the invention provides compounds and methods for the introduction of a predetermined genetic change in a plant cell by introducing an oligodeoxynucleotide encoding the predetermined genetic change into the nucleus of a plant cell.

In preferred embodiments the oligodeoxynucleotide is modified by the attachment of 3' and 5' blocking substituents such as a 3'-3' linked cytosine nucleotide and a 5' linked indocarbocyanine dye. In an alternative embodiment the modification can consist of the replacement of the 3' most and 5' most internucleotide phosphodiester linkage with a non-hydrolyzeable linkage such as a phosphorothioatediester linkage or a phosphoramidate linkage. Alternatively, a 5' linked indocarbocyanine dye and 3' most internucleotide phosphodiester linkage a non-hydrolyzeable linkage can be used in yet a third embodiment.

4. DETAILED DESCRIPTION OF THE INVENTION

The sequence of the SSOMV is based on the same principles as prior art mutational vectors. The sequence of the SSOMV contains two regions that are homologous with the target sequence separated by a region that contains the desired genetic alteration termed the mutator region. The mutator region can have a sequence that is the same length as the sequence that separates the homologous regions in the target sequence, but having a different sequence. Such a mutator region causes a substitution. Alternatively, the homologous regions in the SSOMV can be contiguous to each other, while the regions in the target gene having the same sequence are separated by one, two or more nucleotides. Such a SSOMV causes a deletion from the target gene of the nucleotides that are absent from the SSOMV. Lastly, the sequence of the target gene that is identical to the homologous regions may be adjacent in the target gene but separated by one two or more nucleotides in the sequence of the SSOMV. Such an SSOMV causes an insertion in the sequence of target gene.

The nucleotides of the SSOMV are deoxyribonucleotides that are linked by unmodified phosphodiester bonds except that the 3' terminal and/or 5' terminal

internucleotide linkage or alternatively the two 3' terminal and/or 5' terminal internucleotide linkages can be a phosphorothioate or phosphoramidate. As used herein an internucleotide linkage is the linkage between nucleotides of the SSOMV and does not include the linkage between the 3' end nucleotide or 5' end nucleotide and a blocking substituent, see below.

The length of the SSOMV depends upon the type of cell in which the target gene is located. When the target gene is a chromosomal gene of a mammalian or avian cell the SSOMV is between 25 and 65 nucleotides, preferably between 31 and 59 deoxynucleotides and most preferably between 34 and 48 deoxynucleotides. The total length of the homologous regions is usually the length of the SSOMV less one, two or three nucleotides. A mutator nucleotide can be introduced at more than one position in the SSOMV, which results in more than two homologous regions in the SSOMV. Whether there are two or more homologous regions, the lengths of at least two of the homologous regions should each be at least 8 deoxynucleotides.

For prokaryotic cells, the length of the SSOMV is between 15 and 35 deoxynucleotides. The preferred length of the oligodeoxynucleotide for prokaryotic use depends upon the type of 3' protecting group that is used. When the 3' protecting substituent is a 3'-3' linked deoxycytidine, the oligonucleotide is preferably between about 21 and 28 deoxynucleotides, otherwise the optimal length is between 25 and 35 deoxynucleotides. The lengths of the homology regions are, accordingly, a total length of at least 14 deoxynucleotides and at least two homology regions should each have lengths of at least 7 deoxynucleotides.

For plant cells, the length of the SSOMV is between 21 and 55 deoxynucleotides and the lengths of the homology regions are, accordingly, a total length of at least 20 deoxynucleotides and at least two homology regions should each have lengths of at least 8 deoxynucleotides.

Within these ranges the optimal length of the oligodeoxynucleotide is determined by the GC content, the higher the GC content the shorter the optimal oligodeoxynucleotide. However, a GC content greater than 50% is preferred.

The SSOMV can be used with any type of mammalian, avian or plant cell. It is not important whether the cells are actively replicating or whether the target gene is transcriptionally active. However, when the target gene is located in a bacteria it is important that the bacteria be RecA^+ . Thus, most of the strains of bacteria commonly used in recombinant DNA work are not suitable for use in the present invention because such bacteria are RecA^- in order to reduce the genetic instability of the plasmids cloned therewith.

The SSOMV can be designed to be complementary to either the coding or the non-coding strand of the target gene. When the desired mutation is a substitution of a single base, it is preferred that the mutator nucleotide be a pyrimidine. To the extent that is consistent with achieving the desired functional result it is preferred that both the mutator nucleotide and the targeted nucleotide in the complementary strand be pyrimidines. Particularly preferred are SSOMV that encode transversion mutations, *i.e.*, a C or T mutator nucleotide is mismatched, respectively, with a C or T nucleotide in the complementary strand.

In addition to the oligodeoxynucleotide the SSOMV can contain a 5' blocking substituent that is attached to the 5' terminal carbons through a linker. The chemistry of the linker is not critical other than its length, which should preferably be at least 6 atoms long and that the linker should be flexible.

The chemistry of the 5' blocking substituent for mammalian, avian or plant cells is not critical other than molecular weight which should be less than about 1000 daltons. A variety of non-toxic substituents such as biotin, cholesterol or other steroids or a non-intercalating cationic fluorescent dye can be used. For use in bacterial systems, however, the blocking substituent has a major effect on the

efficiency of the SSOMV and it is preferably a 3,3,3',3'-tetramethyl N,N'-oxyalkyl substituted indocarbocyanine. Particularly preferred as reagents to make SSOMV are the reagents sold as Cy3™ and Cy5™ by Amersham Pharmacia Biotech, Piscataway, NJ, which are blocked phosphoramidites that upon incorporation into an oligonucleotide yield 3,3,3',3'-tetramethyl N,N'-isopropyl substituted indomonocarbocyanine and indodicarbocyanine dyes, respectively. Cy3 is the most preferred. When the indocarbocyanine is N-oxyalkyl substituted it can be conveniently linked to the 5' terminal of the oligodeoxynucleotide through a phosphodiester with a 5' terminal phosphate. The chemistry of the dye linker between the dye and the oligodeoxynucleotide is not critical and is chosen for synthetic convenience. When the commercially available Cy3 phosphoramidite is used as directed the resulting 5' modification consists of a blocking substituent and linker together which are a N-hydroxypropyl, N'-phosphatidylpropyl 3,3,3',3'-tetramethyl indomonocarbocyanine.

In the preferred embodiment the indocarbocyanine dye is tetra substituted at the 3 and 3' positions of the indole rings. Without limitation as to theory these substitutions prevent the dye from being an intercalating dye. The identity of the substituents at these positions are not critical.

The SSOMV can in addition have a 3' blocking substituent. Again the chemistry of the 3' blocking substituent is not critical, other than non-toxicity and molecular weight of less than about 1000, when the target gene is located in other than a bacterial cell. However, when the target gene is located in a bacterial cell the preferred 3' blocking substituent is a so-called inverted nucleotide, *i.e.*, a nucleotide that is linked by an unsubstituted 3'-3' phosphodiester, as is taught by United States patent 5,750,669. In a more preferred embodiment the inverted nucleotide is a thymidine or most preferred a deoxycytidine. For use in bacterial cells, the combination of a Cy3 5' blocking substituent and an inverted deoxycytidine 3' blocking substituent is particularly preferred as the two modifications have a synergistic effect on the efficacy of the SSOMV. The SSOMV with the above recited modifications can be synthesized by conventional solid phase nucleotide synthesis.

The SSOMV can be introduced into the cell containing the target gene by the same techniques that are used to introduce the Kmiec type mutational vectors into mammalian, avian and plant cells. For bacterial cells, a preferred method of introducing the SSOMV is by electroporation.

For use with mammalian and avian cells the preferred method of delivery into the cell is by use of a protective macromolecular carrier. Commercially available liposomal transfecting reagents such Lipofectamine™ and Superfect™ are designed so that the nucleic acid to be transfected is electrostatically adherent to the exposed surface of the liposome. Such carriers are not as preferred as protective macromolecular carriers. Suitable protective macromolecular carriers are disclosed in United States patent application Serial No. 09/108,006, filed June 30, 1998 and in the scientific publication Bandyopadhyay, P., et al., April 1999, J. Biol. Chem. 274, 10163-72, which are each hereby incorporated in their entirety.

A particularly preferred macromolecular carrier is an aqueous-cored lipid vesical or liposome wherein the SSOMV is trapped in the aqueous core. Such vesicals are made by taking a solvent free lipid film and adding an aqueous solution of the SSOMV, followed by vortexing, extrusion or passage through a microfiltration membrane. In one preferred embodiment the lipid constituents are a mixture of dioleoyl phosphatidylcholine/ dioleoyl phosphatidylserine/ galactocerebroside at a ratio of 1:1:0.16. Other carriers include polycations, such as polyethylenimine, having a molecular weight of between 500 daltons and 1.3 Md, with 25 kd being a suitable species and lipid nanospheres, wherein the SSOMV is provided in the form of a lipophilic salt.

When the SSOMV are used to introduce genetic changes in mammalian and avian cells, it is preferred that the macromolecular carrier further comprise a ligand for a cell surface receptor that is internalized. Suitable receptors are the receptors that are internalized by the clathrin-coated pit pathway, such as the asialoglycoprotein receptor, the epidermal growth factor receptor and the transferrin receptor. Also suitable are receptors that are internalized through the caveolar

pathway such as the folic acid receptor. The galactocerebroside is a ligand for the asialoglycoprotein receptor. As used herein an internalizeable receptor is a receptor that is internalized by the clathrin-coated pit pathway or by the caveolar pathway.

The SSOMV can be used for any purpose for which the prior art mutational vectors were employed. Specific uses include the cure of genetic diseases by reversing the disease causing genetic lesion; such diseases includes for example hemophilia, α_1 anti-trypsin deficiency and Crigler-Najjar disease and the other diseases that are taught by patent publication WO 98/49350, which is hereby incorporated by reference in its entirety.

Alternatively, the SSOMV can be used to modify plants for the purposes described in patent publication WO 99/07865, which is hereby incorporated by reference in its entirety. An additional use of SSOMV in plants is the generation of herbicide resistant plants by means that avoid having to introduce a foreign or heterologous gene into a crop plant. Of particular interest is resistance to the herbicide glyphosate (ROUNDUP®). The identity of mutations that confer glyphosate resistance can be found in patent publications WO 99/25853 and WO 97/04103.

Alternatively, the SSOMV can be used to modify bacteria. The use of SSOMV for the genetic manipulation of bacteria is particularly valuable in the fields of antibiotic production and in the construction of specifically attenuated bacteria for the production of vaccines. In both of the above applications it is important that antibiotic resistance genes not remain in the final modified bacteria.

Yet further, the SSOMV can be used in combination with a bacterial artificial chromosome (BAC) to modify a targeted gene from any species that has been cloned into a BAC. A fragment much larger than the targeted gene can be incorporated. The BAC having the cloned targeted gene is placed into a bacterial host and a predetermined genetic change is introduced according to the invention. A BAC subclone having the predetermined genetic change can be identified and the

insert removed for further use. The present invention allows for the predetermined changes to be made without the time and expense attendant with obtaining making PCR fragments and inserting the fragments back into the original gene.

Example 1 Treatment of the Gunn Rat

The Gunn rat contains a mutation in the UDP-glucuronosyltransferase gene, which is the same gene as is mutated in Crigler-Najjar Disease. The Gunn rat mutation Roy-Chowdhury et al., 1991, J. Biol. Chem. 266, 18294; Iyanangi et al., 1989, J. Biol. Chem. 264, 21302. In the Gunn rat there is a mutation at nt 1206 that has deleted a G. A 35 nucleotide SSOMV, termed CN3-35UP, corresponding to the antisense strand was constructed to reverse the mutation and has the following sequence.

5'-ATCATCGGCAGTCATTT **C** CAGGACATTCAGGGTCA-3' (SEQ ID NO: 1)

CN3-35LOW, a second SSOMV that corresponds to the sense strand has the following sequence

5'-TGACCCTGAATGTCCTG **G** AAATGACTGCCGATGAT-3' (SEQ ID NO: 2)

The mutator nucleotide is in bold typeface in the above sequences and in each of the following sequences.

5'-Cy3, 3'-3' dC modified CN3-35UP (2 animals) and CN3-35LOW and unmodified CN3-35UP were formulated in an aqueous cored lipid vesicle having lipid constituents of dioleoyl phosphatidylcholine/ dioleoyl phosphatidylserine/ galactocerebroside at a ratio of 1:1:0.16. Approximately 2.0 ml of 5% dextrose containing 500 µg of the SSOMV was used to hydrate 2 mg of lipid, the vesicles were thereafter extruded to a diameter of 0.5 µm. Encapsulation efficiency was 80%. A positive control group was treated with Kmiec type MV (2 animals) given in an equimolar amount in the same carrier. Rats, weighing 250 gr, were treated on five consecutive days with 300 µg of SSOMV or in the carrier. The resulting serum bilirubin levels were as follows in mg/dl.

R / days post R	0 d	14 d	21 d	26 d	39 d
Unmod-UP	6.3	4.6	5.4	4.2	3.2
Mod-UP	7.9, 6.5	4.1, 3.3	4.9, 5.0	4.2, 3.8	3.6, 3.0
Mod-LOW	6.8	4.3	5.9	4.2	3.5
Kmiec type	6.3, 7.1	4.6, 5.7	4.8, 4.2	5.5, 5.1	4.4, 4.7

The data show that both modified and unmodified SSOMV and that both sense and antisense sequences were at least equivalent and at the longer time points appeared superior to the Kmiec type mutational vectors.

Example 2 Modification of the Human UDP-Glucuronosyltransferase Gene

The following example shows that an unmodified SSOMV in a macromolecular carrier can be used to introduce a specific genetic change in a mammalian cell in an artificial medium at rates that are within a factor of 3 of that seen with Kmiec type DNA/2'OMeRNA mutational vectors. The data further show that modifications as minimal as a single phosphorothioate linkage can result in fully comparable rates.

A group of Amish people have Crigler-Najjar Disease resulting from a C→A substitution at nt 222 of the UDP-Glucuronosyltransferase gene. The mutation results in the conversion of a TAC (Tyr) to a TAA stop codon. A SSOMV designed to introduce the disease causing mutation in a human hepatocellular carcinoma cell line, HuH-7 was designed. A 35 nucleotide SSOMV, designated CNAM3-35UP, or corresponds to the antisense strand and has the following sequence:

5'-GGGTACGTCTTCAAGGT T TAAATGCTCCGTCTCT-3' (SEQ ID NO: 3)

HuH-7 cells at 10^6 / cm² were given 300 µl made in a carrier according to the methods of Example 1 containing CNAM3-35UP, CNAM3-35UP variously modified or an equimolar amount of an 82 nt Kmiec type mutational vector. Cells were harvested and the relevant gene fragment was amplified by PCR, cloned and analyzed by allele specific hybridization according to the methods of Bandyopadhyay *supra*.

The following rates of conversion were observed:

Unmodified SSOMV	6%
5'Cy3 SSOMV	15%
3'-3' dC SSOMV	5%
5'Cy3, 3'-3' dC SSOMV	15%
5' phos'thioate SSOMV	16%
3' phos'thioate	12%
Kmiec type MV	14%.

These data show that in the presence of a macromolecular carrier modified SSOMV were as effective as Kmiec type mutational vectors and that unmodified SSOMV were as effective within a factor of 3.

Example 3 Conversion of Kanamycin Resistance in a BAC

The following example shows that modified SSOMV are more effective than Kmiec DNA/2'OMeRNA mutational vectors in bacterial cells.

A kanamycin resistance gene was inactivated by the insertion of an inframe ATG stop codon. Kanamycin resistance is recovered by converting the third nucleotide to a C, *i.e.*, making a transversion at the third nucleotide.

The sequence of a 41 nt SSOMV that corresponds to the sense strand for the recovery of Kanamycin resistance is as follows

5'-GTGGAGAGGCTATTCGGCTA **C** GACTGGGCACAACAGACAAT-3'

(SEQ ID NO: 4)

To generate pBACKans, a BamHI linker was inserted into the unique SmaI site of pKans, and the resulting 1.3-kb BamHI-HindIII fragment containing the mutant kanamycin gene was inserted into the BamHI/HindIII sites of the BAC cloning vector pBeloBAC11 (Genome Systems, Inc., St. Louis, MO). *Escherichia coli* strains MC1061 and DH10B were transformed with pBACKans, selected on LB chloramphenicol plates, and made electrocompetent.

Forty μ l of electrocompetent cells were electroporated with between 5 and 10 μ g of SSOMV using the following conditions: 25 kV/cm, 200 ohms, 25 microfarads. 1 mL

of SOC was added to cells immediately after electroporation and the culture grown for 1 hour while shaking at 37 C. 4 mL of LB+ chloramphenicol (12.5 µg/mL final) was added and the cultures grown for an additional 2 hours while shaking at 37 C. Appropriate dilutions of the culture were plated on LB-chloramphenicol plates to assess viability and on LB-kanamycin plates to assess conversion. Conversion frequency was calculated by dividing the number of kanamycin resistant colonies/mL by the number of chloramphenicol resistant colonies/mL.

The rate of conversion observed with the 5'Cy3, 3'-3' dC modified 25 nt SSOMV corresponded to about 1 conversion per 100 surviving bacteria.

The relative rates of conversion were

68 nt Kmiec MV w/2'OMe RNA linker	0.04
68 nt Kmiec MV w/DNA linker	0.004
41 nt SSOMV w/3',5' phos'thioate	0.4
35 nt SSOMV w/3',5' phos'thioate	4.0
29 nt SSOMV w/3',5' phos'thioate	0.9
25 nt SSOMV w/3',5' phos'thioate	1.0
41 nt SSOMV w/3'-3' dC,5'Cy3	2.0
35 nt SSOMV w/3'-3' dC,5'Cy3	2.9
35 nt SSOMV w/3'-3' dC,	2.5
35 nt SSOMV w/5'Cy3	2.5
29 nt SSOMV w/3'-3' dC,5'Cy3	4.2
25 nt SSOMV w/3'-3' dC,5'Cy3	42.0
25 nt SSOMV w/3'-3' dC	1.3
25 nt SSOMV w/5'Cy3	1.8
25 nt SSOMV w/3'phos'thioate,5'Cy3	8.4
35 nt SSOMV w/3'phos'thioate,5'Cy3	10.2

These data show that the rate of conversion of the optimal SSOMV was between 10^3 and 10^4 greater than that of the Kmiec type mutational vector.

Example 4 The use of an SSOMV without a Protective Carrier: in a Mammalian Cell- Hygromycin Resistance

This example shows the modification of a mammalian cell using modified SSOMV in the absence of a protective macromolecular carrier. The modified SSOMV were able to introduce the genetic modification at a rate that was between 15 and 30 fold higher than the Kmiec type mutational vectors. This example uses the same gene as in Example 3, however, it is expressed in the HuH-7 cell line.

A clone of HuH7 cells containing a stably integrated copy of the mutant Kanamycin gene in a IRES containing vector (pIRESKan-) were generated under hygromycin selection. Cells were cultured in DMEM high glucose/ 10% FBS containing 100 mg/ml hygromycin to maintain high expression from the integrated construct. Twenty four hours prior to transfection cells were seeded at a density of 1.0×10^6 cells in a 100 mm dish. Two hours prior to transfection the growth medium was replaced with 10 ml of Opti-MEM™. Forty micrograms of oligonucleotide and 40 ml (80 µg) of Lipofectamine™ were diluted in separate tubes containing 200 ml of Opti-MEM pH 8.5. The Lipofectamine is then added to the oligonucleotide, mixed by pipette and incubated at room temperature for 30 minutes before the addition of 3.6 ml of Opti-MEM pH 8.5. The medium is aspirated from the cells and replaced with the 4 ml transfection mixture. The cells are incubated for 2 hours at 37°C before the transfection mix is replaced with standard growth media. Two days post-transfection the cells are split into 2 100 mm dishes in 10 ml media containing 450 mg/ml G418. The G418 containing media is replaced daily for 10 days, then twice a week until colonies are macroscopically visible (16-18 days after transfection). Clones are picked approximately 21 days after transfection and expanded for molecular analysis.

Background rates of the development of hygromycin resistance is about 1 per 10^6 . When Kmiec type mutational vectors were employed there was no increase in the number of resistant colonies. Sequence analysis of one of 5 colonies showed that it had obtained the specific mutation. The mutations in the other 4 colonies could not be identified. When a 41 nt SSOMV w/3'-3' dC,5'Cy3 was used, the rates of development of hygromycin resistant colonies increased by between 15 and 30 fold, i.e., to about 3 per 10^5 . Sequence analysis of these colonies showed that between 100% and 80% of the colonies had the correct genetic change. Experiments with 35 nt SSOMV w/3'-3' dC,5'Cy3 or w/3'phosphorothioate 5'Cy3 or w/two phosphorothioate linkages at each of the 3', 5' ends, each showed rates of development of hygromycin resistance that were about half that of the modified 41 nt SSOMV.

Example 5 The use of an SSOMV without a Protective Carrier in a Mammalian Cell- Tyrosinase

This example shows that in a mammalian cell line an unmodified SSOMV without a protective carrier can be superior to both the 5' Cy3/3'-3' dC modified SSOMV and superior to Kmiec type DNA/2'OMe RNA mutational vectors.

These experiments use Melan-c a murine melanocyte cell line having a C→G mutation at codon 82 of the tyrosinase gene, which creates an inframe stop.

Bennett, et al., 1989, Development 105, 379-385. A 35 nt SSOMV which corresponds to the coding sequence was designed and has the following sequence: 5'-CCCCAAATCCAAACTTA C AGTTTCCGCGAGTTGAAA-3' (SEQ ID NO: 5)

Melan-c cells were cultured in RPMI medium containing 10% fetal bovine serum, 100 nM phorbol 12-myristate 13-acetate (PMA) and 0.1 mM b-mercaptoethanol (Gibco, Bethesda, MD). Two days prior to transfection, cells were seeded at a density of 0.5-1.5 x10⁵ cells/well in a 6 six-well plate and refed with fresh medium 24 hours prior to transfection. Five to ten micrograms (220-440 nM) of the oligonucleotides, were incubated with 6-9 µg of Superfectin™ in 0.1 ml of TE (10 mM TRIS pH 7.5, 1 mM EDTA) for 30 min at room temperature. The transfection mixture was added to the cells containing 0.9 ml of DMEM high glucose growth media containing 10% serum and 100 nM PMA. After 6-18 hours, cells were washed with phosphate-buffered saline and fed with 2 ml of the DMEM media. Cells were monitored for a change in pigmentation by microscopy. The number of conversion events was determined by counting the number of pigmented cells or cell clusters 5 to 8 days after transfection.

The rates of albino→wild type (pigmented) conversion per 10⁵ cells as follows:

Kmiec type MV	1
unmodified SSOMV	5
SSOMV w/3',5' phos'thioate	6
SSOMV w/3'-3' dC	2
SSOMV w/5' Cy3	3
SSOMV w/3'-3' dC, 5' Cy3	1

Example 6 The Use of a Modified SSOMV in Plants

This example concerns the use of a SSOMV to introduce a Ser→Asn mutation at position 653 of the *Arabidopsis thaliana* acetohydroxyacid synthase (also known as

acetolactate synthase). The mutation requires that an AGT codon be converted to a AAT codon and introduces resistance to imidazoline herbicides as well as sulfonyl urea herbicides.

A 25 nt SSOMV and a 35 nt SSOMV were synthesized having 3'-3' dC and 5' Cy3 modifications and had the following sequences:

5'-CGATCCCGA A TGGTGGCACTTT-3' (SEQ ID NO: 6)

5'-GTTGCCGATCCCGA A TGGTGGCACTTTCAACG-3' (SEQ ID NO: 7)

A disaggregated *A. thaliana* cell population was prepared plated at 10^6 per plate and subjected to biolistic introduction of the SSOMV or a Kmiec type MV having the same sequence. Control plates using a plasmid determined that the efficiency of the biolistic system is about one delivery per 200 cells plated. After two months selection with 10 μ M Imazaquin™ each of the biolistically treated cell populations showed a background corrected rate of Imazaquin resistance of about 1 per 10^3 cells into which the mutational vectors had been successfully introduced.

Example 7 Preparation of Folate-conjugated PEI

This example describes the preparation of folate-conjugated PEI which is suitable to use as a macromolecular carrier in the invention.

Folic acid (4.4 mg, 10 μ mole) in sodium phosphate buffer (1.5 mL, 133 mM, pH 4.5) was treated with 200 μ L pyridine and 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride (EDC, 15.5. mg, 98 μ mol) and incubated at room temperature for 1 hr. The activated folate solution (1.7 mL) was added to an aqueous solution of polyethyleneimine (25 kDa, 24.55 mg/mL; 1.02 mL) and incubated for 3 days at RT with gentle agitation. The conjugated polyethyleneimine was purified by dialysis against water through a 12 kDa MW cutoff membrane. The product was positive for amines by the ninhydrin assay and folate by UV absorbance with maxima at 259, 289 and 368 nm.

Coupling was about 1-2 folate moieties per 1000 amines which is equivalent to 1-2 folate per PEI molecule.

We claim:

1. A composition for making a predetermined genetic change in a targeted chromosomal gene of a mammalian or avian cell, comprising:
 - a. a single-stranded oligodeoxynucleotide having a 3' end nucleotide, a 5' end nucleotide, and having at least 25 deoxynucleotides and not more than 65 deoxynucleotides and having a sequence comprising at least two regions of at least 8 deoxynucleotides that are each, respectively identical to two regions of the targeted chromosomal gene, which regions are separated by at least one nucleotide in the sequence of the targeted chromosomal gene or in the sequence of the oligodeoxynucleotide or both, and which regions together are at least 24 nucleotides in length; and
 - b. a macromolecular carrier selected from the group consisting of
 - i. an aqueous-cored lipid vesicle, wherein the aqueous core contains the single-stranded oligonucleotide
 - ii. a lipid nanosphere, which comprises a lipophilic salt of the single-stranded oligonucleotide, and
 - iii. a polycation having an average molecular weight of between 500 daltons and 1.3 Md wherein the polycation forms a salt with the oligonucleobase.
2. The composition of claim 1 in which the length of the single-stranded oligonucleotide is at least 31 deoxynucleotides and not more than 59 deoxynucleotides.
3. A method of obtaining a mammalian or avian cell that contains a predetermined genetic change in a target gene which comprises:
 - a. providing a population of mammalian or avian cells in a culture media;
 - b. adding the composition of claim 2 to the culture media; and
 - c. identifying a cell of the population having the predetermined genetic change.

4. The method of claim 3, which further comprises isolating the identified cell.
5. The composition of claim 1, in which the macromolecular carrier further comprises a ligand for an internalizeable receptor of the mammalian cell that is affixed to the surface of the macromolecular carrier.
6. A method of making a predetermined genetic change in a tissue of a subject mammal which comprises:
 - a. administering to the subject mammal the composition of claim 5 in a pharmaceutically acceptable carrier; and
 - b. detecting the presence of the predetermined genetic change in the cells of the tissue of the subject mammal.
7. The method of claim 6, wherein the subject mammal is a human having a genetic lesion that is reversed by the predetermined genetic change which comprises administering an amount of the composition which is effective to ameliorate the effects of the genetic lesion.
8. The method of claim 6, wherein the tissue is the liver.
9. The composition of claim 5, in which the receptor is selected from the group consisting of the asialoglycoprotein receptor, the transferrin receptor and the epidermal growth factor receptor.
10. The composition of claim 5, in which the receptor is the folic acid receptor.
11. The composition of claim 1, in which the internucleotide linkage attached to the 3' end nucleotide is a phosphorothioate linkage.
12. The composition of claim 1, in which the internucleotide linkage attached to the 5' end nucleotide is a phosphorothioate linkage.

13. The composition of claim 1, in which the 5' hydroxyl of the 5' end nucleotide is attached to a 5' blocking substituent.
14. The composition of claim 13, in which the 5' blocking substituent is a N'-hydroxyalkyl substituted 3,3,3',3'-tetra substituted indocarbocyanine dye, which is attached to the 5' hydroxyl through a linker.
15. The composition of claim 14, in which the indocarbocyanine dye and linker together are a N-hydroxypropyl, N'-phosphatidylpropyl 3,3,3',3'-tetramethyl indomonocarbocyanine.
16. The composition of claim 14, in which the internucleotide linkage attached to the 3' end nucleotide is a phosphorothioate linkage.
17. The composition of claim 1, in which the 3' hydroxy of the 3' end nucleotide is attached to a 3' blocking substituent.
18. The composition of claim 17, in which the 3' blocking substituent is a blocking nucleotide that is 3'-3' linked to the 3' hydroxy of the 3' end nucleotide.
19. A compound for making a predetermined genetic change in a targeted chromosomal gene of a mammalian or avian cell, comprising: a single-stranded oligodeoxynucleotide having a 3' end nucleotide, a 5' end nucleotide, and having at least 25 deoxynucleotides and not more than 65 deoxynucleotides and having a sequence comprising at least two regions of at least 8 deoxynucleotides that are each, respectively, identical to two regions of the targeted chromosomal gene, which regions are separated by at least one nucleotide in the sequence of the targeted chromosomal gene or in the sequence of the oligodeoxynucleotide or both, and which regions together are at least 24 nucleotides in length.

20. A method of obtaining a mammalian or avian cell that contains a predetermined genetic change in a target gene which comprises:
- providing a population of mammalian or avian cells in a culture media;
 - adding the compound of claim 19 to the culture media; and
 - identifying a cell of the population having the predetermined genetic change.
21. The compound of claim 19, in which the internucleotide linkage attached to the 3' end nucleotide is a phosphorothioate linkage.
22. The compound of claim 21, in which the internucleotide linkage attached to the 5' end nucleotide is a phosphorothioate linkage.
23. A method of obtaining a mammalian or avian cell that contains a predetermined genetic change in a target gene which comprises:
- providing a population of mammalian or avian cells in a culture media;
 - adding the composition of claim 22 to the culture media; and
 - identifying a cell of the population having the predetermined genetic change.
24. The compound of claim 21, in which an N'- hydroxyalkyl substituted 3,3,3',3'-tetra substituted indocarbocyanine dye is attached to the 5' hydroxyl of the 5' end nucleotide through a linker.
25. The compound of claim 19, in which the internucleotide linkage attached to the 5' end nucleotide is a phosphorothioate linkage.
26. The compound of claim 25, in which the internucleotide linkage attached to the 3' end nucleotide is a phosphorothioate linkage or in which a deoxycytidine or thymidine nucleotide is 3'-3' linked to the 3' hydroxy of the 3' end nucleotide or both.

27. A compound for making a predetermined genetic change in a targeted gene in a bacterial cell, comprising:
- a. a single-stranded oligonucleotide having a 3' end nucleotide, a 5' end nucleotide, and having at least 15 deoxynucleotides and not more than 35 deoxynucleotides and having a sequence comprising at least two regions of at least 7 deoxynucleotides that are each, respectively, identical to two regions of the targeted gene, which regions are separated by at least one nucleotide in the sequence of the targeted gene or in the sequence of the single-stranded oligonucleotide or both, and which regions together are at least 14 nucleotides in length;
 - b. a 5' modification wherein the internucleotide linkage attached to the 5' end nucleotide is a phosphorothioate linkage or wherein a N'-hydroxyalkyl substituted 3,3,3',3'-tetra substituted indocarbocyanine dye is attached through a linker to the 5' hydroxyl of the 5' end nucleotide; and
 - c. a 3' modification wherein the internucleotide linkage attached to the 3' end nucleotide is a phosphorothioate linkage or wherein a deoxycytidine or thymidine nucleotide is 3'-3' linked to the 3' hydroxy of the 3' end nucleotide or both.
28. The compound of claim 27, in which the 5' modification comprises a N-hydroxypropyl, N'-phosphatidylpropyl 3,3,3',3'-tetramethyl indomonocarbocyanine.
29. The compound of claim 27, in which the 3' modification consists of a 3'-3' linked deoxycytidine.
30. The compound of claim 27, in which the 3' modification consists of a phosphorothioate internucleotide linkage attached to the 3' end nucleotide.
31. A compound for making a predetermined genetic change in a targeted gene in a plant cell, comprising a single-stranded oligonucleotide having a 3' end

nucleotide, a 5' end nucleotide, and having at least 21 deoxynucleotides and not more than 55 deoxynucleotides and having a sequence comprising at least two regions of at least 8 nucleotides that are each, respectively identical to two regions of the targeted gene, which regions are separated by at least one nucleotide in the sequence of the targeted gene or in the sequence of the single-stranded oligonucleotide or both, and which regions together are at least 20 nucleotides in length.

32. A method of obtaining a plant cell that contains a predetermined genetic change in a target gene which comprises:
 - a. introducing the compound of claim 31 into a population of plant cells; and
 - b. identifying a cell of the population having the predetermined genetic change.
33. The method of claim 32, which further comprises isolating the identified cell.
34. The compound of claim 31, in which the 5' hydroxyl of the 5' end nucleotide is attached to a 5' blocking substituent.
35. The compound of claim 34, in which the 3' hydroxyl of the 3' end nucleotide is attached to a 3' blocking substituent.
36. The compound of claim 35, in which
 - a. the 5' blocking substituent is a N'- hydroxyalkyl substituted 3,3,3',3'-tetra substituted indocarbocyanine dye, which is attached through a linker to the 5' hydroxyl of the 5' end nucleotide; and
 - b. the 3' blocking substituent is a blocking nucleotide that is 3'-3' linked to the 3' hydroxyl of the 3' end nucleotide.
37. The compound of claim 36, in which the single stranded oligonucleotide is at least 25 nucleotides and not more than 35 nucleotides in length.

38. The compound of claim 36, in which the blocking nucleotide is a deoxycytidine or thymidine.
39. The compound of claim 36, in which the indocarbocyanine dye and linker together are a N-hydroxypropyl, N'-phosphatidylpropyl 3,3,3',3'-tetramethyl indomonocarbocyanine.
40. The compound of claim 31, in which the internucleotide linkage attached to the 3' end nucleotide is a phosphorothioate linkage.
41. The compound of claim 31, in which the internucleotide linkage attached to the 5' end nucleotide is a phosphorothioate linkage.
42. The compound of claim 31, in which
 - a. the 5' hydroxyl of the 5' end nucleotide is attached to a 5' blocking substituent; and
 - b. the internucleotide linkage attached to the 3' end nucleotide is a phosphorothioate linkage.
43. The compound of claim 42, in which the 5' blocking substituent is a N'-hydroxyalkyl substituted 3,3,3',3'-tetra substituted indocarbocyanine dye, which is attached through a linker to the 5' hydroxyl of the 5' end nucleotide;

ABSTRACT

The invention concerns the introduction of predetermined genetic changes in target genes of a living cell by introducing an oligodeoxynucleotide encoding the predetermined change. The oligodeoxynucleotides are effective in mammalian, avian, plant and bacterial cells. Specific end modifications that greatly increase the effectiveness of the oligodeoxynucleotides in bacteria are described. Surprisingly, unmodified oligodeoxynucleotides can be as effective in mammalian cells, including *in vivo* hepatocytes, as the modified nucleotides and can be as effective or more effective than chimeric oligonucleotides that consist of a mixture of deoxynucleotides and 2'-O-methyl ribonucleotides.

SEQUENCE LISTING

<110> Richard A. Metz
Bruce L. Frank
Debra M. Walther

<120> Single-Stranded Oligodeoxynucleotide
Mutational Vectors

<130> 7991-046-999

<160> 7

<170> FastSEQ for Windows Version 3.0

<210> 1

<211> 35

<212> DNA

<213> Artificial Sequence

<220>

<223> Mutational Vector

<400> 1

atcatcggca gtcatttcca ggacattcag ggtca

35

<210> 2

<211> 35

<212> DNA

<213> Artificial Sequence

<400> 2

tgaccctgaa tgtcctggaa atgactgccg atgat

35

<210> 3

<211> 35

<212> DNA

<213> Artificial Sequence

<400> 3

gggtacgtct tcaaggttta aaatgctccg tctct

35

<210> 4

<211> 41

<212> DNA

<213> Artificial Sequence

<400> 4

gtggagaggc tattcggcta cgactgggca caacagacaa t

41

<210> 5

<211> 35

<212> DNA

<213> Artificial Sequence

<400> 5
ccccaaatcc aaacttacag tttccgcagt tgaaa

35

<210> 6
<211> 25
<212> DNA
<213> Artificial Sequence

<400> 6
cgatgatccc gaatggtggc acttt

25

<210> 7
<211> 35
<212> DNA
<213> Artificial Sequence

<400> 7
gttgccgatg atcccgaatg gtggcacttt caacg

35

Pending Application 09/129,298

Filed on 08/05/1998

Inventors: Kipp *et al.*

Our file: 7991-023-999

ATTORNEY DOCKET NUMBER: 7991-086-99

SERIAL NUMBER: 09/685,403

REFERENCE: **BD**

THE USE OF MIXED DUPLEX OLIGONUCLEOTIDES TO EFFECT LOCALIZED GENETIC CHANGES IN PLANTS

1. FIELD OF THE INVENTION

The field of the present invention relates to methods for the improvement of existing lines of plants and to the development of new lines having desired traits. The previously available methods of obtaining genetically altered plants by recombinant DNA technology enabled the introduction of preconstructed exogenous genes in random, atopic positions, so-called transgenes. In contrast the present invention allows the skilled practitioner to make a specific alteration of a specific pre-existing gene of a plant. The invention utilizes duplex oligonucleotides having a mixture of RNA-like nucleotides and DNA-like nucleotides to effect the alterations, hereafter "mixed duplex oligonucleotides" or MDON.

2. BACKGROUND TO THE INVENTION

2.1 MDON and Their Use to Effect Specific Genetic Alterations

Mixed duplex oligonucleotides (MDON) and their use to effect genetic changes in eukaryotic cells are described in United States patent No. 5,565,350 to Kmiec (Kmiec I). Kmiec I discloses *inter alia* MDON having two strands, in which a first strand contains two segments of at least 8 RNA-like nucleotides that are separated by a third segment of from 4 to about 50 DNA-like nucleotides, termed an "interposed DNA segment." The nucleotides of the first strand are base paired to DNA-like nucleotides of a second strand. The first and second strands are additionally linked by a segment of single stranded nucleotides so that the first and second strands are parts of a single oligonucleotide chain. Kmiec I further teaches a method for introducing specific genetic alterations into a target gene. According to Kmiec I, the sequences of the RNA segments are selected to be homologous, i.e., identical, to the sequence of a first and a second fragment of the target gene. The sequence of the interposed DNA segment is homologous with the sequence of the target gene between the first and second fragment except for a region of difference, termed the "heterologous region." The heterologous region can effect an insertion or deletion, or can contain one or

more bases that are mismatched with the sequence of target gene so as to effect a substitution. According to Kmiec I, the sequence of the target gene is altered as directed by the heterologous region, such that the target gene becomes homologous with the sequence of the MDON. Kmiec I specifically teaches that ribose and 2'-O-methylribose, i.e., 2'-methoxyribose, containing nucleotides can be used in MDON and that naturally-occurring deoxyribose-containing nucleotides can be used as DNA-like nucleotides.

United States patent application Serial No. 08\664,487, filed June 17, 1996, now U.S. patent No. 5,731,181 (Kmiec II) does specifically disclose the use of MDON to effect genetic changes in plant cells and discloses further examples of analogs and derivatives of RNA-like and DNA-like nucleotides that can be used to effect genetic changes in specific target genes.

Scientific publications disclosing uses of MDON having interposed DNA segments include Yoon, et al., 1996, *Proc. Natl. Acad. Sci.* 93:2071-2076 and Cole-Straus, A. et al., 1996, *SCIENCE* 273 :1386-1389. The scientific publications disclose that rates of mutation as high as about one cell in ten can be obtained using liposomal mediated delivery. However, the scientific publications do not disclose that MDON can be used to make genetic changes in plant cells.

The present specification uses the term MDON, which should be understood to be synonymous with the terms "chimeric mutation vector," "chimeric repair vector" and "chimeraplast" which are used elsewhere.

2.2 Transgenic Plant Cells and the Generation of Plants from Transgenic Plant Cells

Of the techniques taught by Kmiec I and II for delivery of MDON into the target cell, the technique that is most applicable for use with plant cells is the electroporation of protoplasts. The regeneration of fertile plants from protoplast cultures has been reported for certain species of dicotyledonous plants, e.g., *Nicotiana tabacum* (tobacco), United States Patent 5,231,019 and Fromm, M.E., et al., 1988, *Nature* 312, 791, and soybean variety *Glycine max*, WO 92/17598 to Widholm, J.M. However, despite the reports of isolated successes using non-transformed cells, Prioli, L.M., et al., *Bio/Technology* 7, 589, Shillito, R.D., et al., 1989, *Bio/Technology* 7, 581, the regeneration of fertile monocotyledonous plants from transformed protoplast

cultures is not regarded as obtainable with application of routine skill. Frequently, transformed protoplasts of monocotyledonous plants result in non-regenerable tissue or, if the tissue is regenerated the resultant plant is not fertile.

Other techniques to obtain transformed plant cells by introducing kilobase-sized plasmid DNA into plant cells having intact or partially intact cell walls have been developed. United States patent No. 4,945,050, No. 5,100,792 and No. 5,204,253 concern the delivery of plasmids into intact plant cells by adhering the plasmid to a microparticle that is ballistically propelled across the cell wall, hereafter "biolistically transformed" cell. For example U.S. patent No. 5,489,520 describes the regeneration of a fertile maize plant from a biolistically transformed cell. Other techniques for the introduction of plasmid DNA into suspensions of plant cells having intact cell walls include the use of silicon carbide fibers to pierce the cell wall, see U.S. patent No. 5,302,523 to Coffee R., and Dunwell, J.M.

A technique that allows for the electroporation of maize cells having a complex cell wall is reported in U.S. patent No. 5,384,253 to Krzyzek, Laursen and P.C. Anderson. The technique uses a combination of the enzymes endopectin lyase (E.C. 3.2.1.15) and endopolygalacturonase (E.C. 4.2.2.3) to generate transformation competent cells that can be more readily regenerated into fertile plants than true protoplasts. However, the technique is reported to be useful only for F1 cell lines from the cross of line A188 x line B73.

3. SUMMARY OF THE INVENTION

The present invention provides new methods of use of the MDON that are particularly suitable for use in such plant cells.

Thus one aspect of the invention is techniques to adhere MDON to particles which can be projected through the cell wall to release the MDON within the cell in order to cause a mutation in a target gene of the plant cell. The mutations that can be introduced by this technique are mutations that confer a growth advantage to the mutated cells under appropriate conditions and mutations that cause a phenotype that can be detected by visual inspection. Such mutations are termed "selectable mutations."

In a further embodiment the invention encompasses a method of introducing a

mutation other than a selectable mutation into a target gene of a plant cell by a process which includes the steps of introducing a mixture of a first MDON that introduces a selectable mutation in the plant cell and a second MDON that causes the non-selectable mutation.

The invention further encompasses the culture of the cells mutated according to the foregoing embodiments of the invention so as to obtain a plant that produces seeds, henceforth a "fertile plant," and the production of seeds and additional plants from such a fertile plant.

The invention further encompasses fertile plants having novel characteristics which can be produced by the methods of the invention.

4. DETAILED DESCRIPTION OF THE INVENTION

4.1 Recombinagenic Oligonucleobases and Mixed Duplex Oligonucleotides

The invention can be practiced with MDON having the conformations and chemistries described in Kmiec I or in Kmiec II, which are hereby incorporated by reference. The MDON of Kmiec I and/or Kmiec II contain two complementary strands, one of which contains at least one segment of RNA-type nucleotides (an "RNA segment") that are base paired to DNA-type nucleotides of the other strand.

Kmiec II discloses that purine and pyrimidine base-containing non-nucleotides can be substituted for nucleotides. Commonly assigned U.S. patent applications Serial No. 09/078,063, filed May 12, 1998, and Serial No. 09/078,064, filed May 12, 1998, which are each hereby incorporated in their entirety, disclose additional molecules that can be used for the present invention. The term "recombinagenic oligonucleobase" is used herein to denote the molecules that can be used in the present invention. Recombinagenic oligonucleobases include MDON, non-nucleotide containing molecules taught in Kmiec II and the molecules taught in the above noted commonly assigned patent applications.

In a preferred embodiment the RNA-type nucleotides of the MDON are made Rnase resistant by having replacing the 2'-hydroxyl with a fluoro, chloro or bromo functionality or by placing a substituent on the 2'-O. Suitable substituents include the

substituents taught by the Kmiec II, C₁₋₆ alkane. Alternative substituents include the substituents taught by U.S. Patent No. 5,334,711 (Sproat) and the substituents taught by patent publications EP 629 387 and EP 679 657 (collectively, the Martin Applications), which are hereby incorporated by reference. As used herein a 2'-fluoro, chloro or bromo derivative of a ribonucleotide or a ribonucleotide having a 2'-OH substituted with a substituent described in the Martin Applications or Sproat is termed a "2'-Substituted Ribonucleotide." As used herein the term "RNA-type nucleotide" means a 2'-hydroxyl or 2'-Substituted Nucleotide that is linked to other nucleotides of a MDON by an unsubstituted phosphodiester linkage or any of the non-natural linkages taught by Kmiec I or Kmiec II. As used herein the term "deoxyribo-type nucleotide" means a nucleotide having a 2'-H, which can be linked to other nucleotides of a MDON by an unsubstituted phosphodiester linkage or any of the non-natural linkages taught by Kmiec I or Kmiec II.

A particular embodiment of the invention comprises MDON that are linked solely by unsubstituted phosphodiester bonds. Alternatively embodiments comprise linkage by substituted phosphodiesters, phosphodiester derivatives and non-phosphorus-based linkages as taught by Kmiec II. A further particular embodiment comprises MDON wherein each RNA-type nucleotide is a 2'-Substituted Nucleotide. Particular preferred embodiments of 2'-Substituted Ribonucleotides are 2'-fluoro, 2'-methoxy, 2'-propyloxy, 2'-allyloxy, 2'-hydroxylethyloxy, 2'-methoxyethyloxy, 2'-fluoropropyloxy and 2'-trifluoropropyloxy substituted ribonucleotides. In more preferred embodiments of 2'-Substituted Ribonucleotides are 2'-fluoro, 2'-methoxy, 2'-methoxyethyloxy, and 2'-allyloxy substituted nucleotides. In one embodiment the MDON oligomer is linked by unsubstituted phosphodiester bonds.

Although MDON having only a single type of 2'-substituted RNA-type nucleotide are more conveniently synthesized, the invention can be practiced with MDON having two or more types of RNA-type nucleotides. The function of an RNA segment may not be affected by an interruption caused by the introduction of a deoxynucleotide between two RNA-type trinucleotides, accordingly, the term RNA segment encompasses such an "interrupted RNA segment." An uninterrupted RNA segment is termed a contiguous RNA segment. In an alternative embodiment an RNA segment can contain alternating RNase-resistant and unsubstituted 2'-OH nucleotides.

The MDON of the invention preferably have fewer than 100 nucleotides and more preferably fewer than 85 nucleotides, but more than 50 nucleotides. The first and second strands are Watson-Crick base paired. In one embodiment the strands of the MDON are covalently bonded by a linker, such as a single stranded hexa, penta or tetranucleotide so that the first and second strands are segments of a single oligonucleotide chain having a single 3' and a single 5' end. The 3' and 5' ends can be protected by the addition of a "hairpin cap" whereby the 3' and 5' terminal nucleotides are Watson-Crick paired to adjacent nucleotides. A second hairpin cap can, additionally, be placed at the junction between the first and second strands distant from the 3' and 5' ends, so that the Watson-Crick pairing between the first and second strands is stabilized.

The first and second strands contain two regions that are homologous with two fragments of the target gene, i.e., have the same sequence as the target gene. A homologous region contains the nucleotides of an RNA segment and may contain one or more DNA-type nucleotides of connecting DNA segment and may also contain DNA-type nucleotides that are not within the intervening DNA segment. The two regions of homology are separated by, and each is adjacent to, a region having a sequence that differs from the sequence of the target gene, termed a "heterologous region." The heterologous region can contain one, two or three mismatched nucleotides. The mismatched nucleotides can be contiguous or alternatively can be separated by one or two nucleotides that are homologous with the target gene. Alternatively, the heterologous region can also contain an insertion or one, two, three or of five or fewer nucleotides. Alternatively, the sequence of the MDON may differ from the sequence of the target gene only by the deletion of one, two, three, or five or fewer nucleotides from the MDON. The length and position of the heterologous region is, in this case, deemed to be the length of the deletion, even though no nucleotides of the MDON are within the heterologous region. The distance between the fragments of the target gene that are complementary to the two homologous regions is identically the length of the heterologous region when a substitution or substitutions is intended. When the heterologous region contains an insertion, the homologous regions are thereby separated in the MDON farther than their complementary homologous fragments are in the gene, and the converse is applicable

when the heterologous region encodes a deletion.

The RNA segments of the MDON are each a part of a homologous region, i.e., a region that is identical in sequence to a fragment of the target gene, which segments together preferably contain at least 13 RNA-type nucleotides and preferably from 16 to 25 RNA-type nucleotides or yet more preferably 18-22 RNA-type nucleotides or most preferably 20 nucleotides. In one embodiment, RNA segments of the homology regions are separated by and adjacent to, i.e., "connected by" an intervening DNA segment. In one embodiment, each nucleotide of the heterologous region is a nucleotide of the intervening DNA segment. An intervening DNA segment that contains the heterologous region of a MDON is termed a "mutator segment."

Commonly assigned U.S. patent application Serial No. 09/078,063, filed May 12, 1998, and Serial No. 09/078,064, filed May 12, 1998, disclose a type of duplex recombinagenic oligonucleobase in which a strand has a sequence that is identical to that of the target gene and only the sequence of the "complementary" strand contains a heterologous region. This configuration results in one or more mismatched bases or a "heteroduplex" structure. The heterologous region of the heteroduplex recombinagenic oligonucleobases that are useful in the present invention is located in the strand that contains the deoxynucleotides. In one embodiment, the heterologous region is located on the strand that contains the 5' terminal nucleotide.

4.2 The Location and Type of Mutation Introduced by a MDON

Frequently, the design of the MDON for use in plant cells must be modified from the designs taught in Kmiec I and II. In mammalian and yeast cells, the genetic alteration introduced by a MDON that differs from the target gene at one position is the replacement of the nucleotide in the target gene at the mismatched position by a nucleotide complementary to the nucleotide of the MDON at the mismatched position. By contrast, in plant cells there can be an alteration of the nucleotide one base 5' to the mismatched position on the strand that is complementary to the strand that contains the DNA mutator segment. The nucleotide of the target gene is replaced by a nucleotide complementary to the nucleotide of the DNA mutator segment at the mismatched position. Consequently, the mutated target gene differs from the MDON at two positions.

The mutations introduced into the target gene by a MDON are located between the regions of the target gene that are homologous with the ribonucleotide portion of the homology regions of the MDON, henceforth the "RNA segments." The specific mutation that is introduced depends upon the sequence of the heterologous region. An insertion or deletion in the target gene can be introduced by using a heterologous region that contains an insertion or deletion, respectively. A substitution in the target gene can be obtained by using a MDON having a mismatch in the heterologous region of the MDON. In the most frequent embodiments, the mismatch will convert the existing base of the target gene into the base that is complementary to the mismatched base of the MDON. The location of the substitution in the target gene can be either at the position that corresponds to the mismatch or, more frequently, the substitution will be located at the position on the target strand immediately 5' to the position of the mismatch, i.e., complementary to the position of the MDON immediately 3' of the mismatched base of the MDON.

The relative frequency of each location of the mismatch-caused substitution will be characteristic of a given gene and cell type. Thus, those skilled in the art will appreciate that a preliminary study to determine the location of substitutions in the gene of particular interest is generally indicated, when the location of the substitution is critical to the practice of the invention.

4.3 The Delivery of MDON by Microcarriers and Microfibers

The use of metallic microcarriers (microspheres) for introducing large fragments of DNA into plant cells having cellulose cell walls by projectile penetration is well known to those skilled in the relevant art (henceforth biolistic delivery). United States patents No. 4,945,050, No. 5,100,792 and No. 5,204,253 concern general techniques for selecting microcarriers and devices for projecting them.

The conditions that are used to adhere DNA fragments to the microcarriers are not suitable for the use of MDON. The invention provides techniques for adhering sufficient amounts of MDON to the microcarrier so that biolistic delivery can be employed. In a suitable technique, ice cold microcarriers (60 mg/ml), MDON (60 mg/ml) 2.5 M CaCl_2 and 0.1 M spermidine are added in that order; the mixture gently agitated, e.g., by vortexing, for 10 min and allowed to stand at room temperature for

10 min, whereupon the microcarriers are diluted in 5 volumes of ethanol, centrifuged and resuspended in 100% ethanol. Good results can be obtained with a concentration in the adhering solution of 8-10 $\mu\text{g}/\mu\text{l}$ microcarriers, 14-17 $\mu\text{g}/\text{ml}$ MDON, 1.1-1.4 M CaCl_2 and 18-22 mM spermidine. Optimal results were observed under the conditions of 8 $\mu\text{g}/\mu\text{l}$ microcarriers, 16.5 $\mu\text{g}/\text{ml}$ MDON, 1.3 M CaCl_2 and 21 mM spermidine.

MDON can also be introduced into plant cells for the practice of the invention using microfibers to penetrate the cell wall and cell membrane. U.S. Patent No. 5,302,523 to Coffee et al. describes the use of $30 \times 0.5 \mu\text{m}$ and $10 \times 0.3 \mu\text{m}$ silicon carbide fibers to facilitate transformation of suspension maize cultures of Black Mexican Sweet. Any mechanical technique that can be used to introduce DNA for transformation of a plant cell using microfibers can be used to deliver MDON for transmutation.

A suitable technique for microfiber delivery of MDON is as follows. Sterile microfibers (2 μg) are suspended in 150 μl of plant culture medium containing about 10 μg of MDON. A suspension culture is allowed to settle and equal volumes of packed cells and the sterile fiber/MDON suspension are vortexed for 10 minutes and plated. Selective media are applied immediately or with a delay of up to about 120 hours as is appropriate for the particular trait.

The techniques that can be used to deliver MDON to transmute nuclear genes can also be used to cause transmutation of the genes of a plastid of a plant cell. Plastid transformation of higher plants by biolistic delivery of a plasmid followed by an illegitimate recombinatorial insertion of the plasmid is well known to those skilled in the art. Svab, Z., et al., 1990, Proc. Natl. Acad. Sci. 87, 8526-8530. The initial experiments showed rates of transformation that were between 10-fold and 100-fold less than the rate of nuclear transformation. Subsequent experiments showed that rates of plasmid transformation comparable to the rate of nuclear transformation could be achieved by use of a dominant selectable trait such as a bacterial aminoglycoside 3'-adenosyltransferase gene, which confers spectinomycin resistance. Svab, Z., & Maliga, P., 1993, Proc. Natl. Acad. Sci. 90, 913-917.

According to the invention MDON for the transmutation of plastid genes can be introduced into plastids by the same techniques as above. When the mutation

desired to be introduced is a selectable mutation the MDON can be used alone. When the desired mutation is non-selectable the relevant MDON can be introduced along with a MDON that introduces a selectable plastid mutation, e.g., a mutation in the psbA gene that confers triazine resistance, or in combination with a linear or circular plasmid that confers a selectable trait.

The foregoing techniques can be adapted for use with recombinagenic oligonucleobases other than MDON.

4.4 Protoplast Electroporation

In an alternative embodiment the recombinagenic oligonucleobase can be delivered to the plant cell by electroporation of a protoplast derived from a plant part. The protoplasts are formed by enzymatic treatment of a plant part, particularly a leaf, according to techniques well known to those skilled in the art. See, e.g., Gallois et al., 1996, in *Methods in Molecular Biology* 55, 89-107 (Humana Press, Totowa, NJ). The protoplasts need not be cultured in growth media prior to electroporation.

Suitable conditions for electroporation are 3×10^5 protoplasts in a total volume of 0.3 ml with a concentration of MDON of between 0.6 - 4 $\mu\text{g/mL}$.

4.5 The Introduction of Mutations

The invention can be used to effect genetic changes, herein "transmutate," in plant cells. In an embodiment the plant cells have cell walls, i.e., are other than protoplasts.

The use of MDON to transmutate plant cells can be facilitated by co-introducing a trait that allows for the ready differentiation and separation of cells (hereafter "selection") into which MDON have been introduced from those that have not. In one embodiment of the invention the selection is performed by forming a mixture of MDON and a plasmid that causes the transient expression of a gene that confers a selectable trait, i.e., one that permits survival under certain conditions, e.g., a kanamycin resistance gene. Under these circumstances elimination of cells lacking the selectable trait removes the cells into which MDON were not introduced. The use of a transient expression plasmid to introduce the selectable trait allows for the successive introduction of multiple genetic changes into a plant cell by repeatedly

using a single standardized selection protocol.

In an alternative embodiment transmutation can be used to introduce a selectable trait. A mixture of a first MDON that causes a selectable mutation in a first target gene and a second MDON that causes a non-selectable mutation in a second target gene is prepared. According to the invention, at least about 1% of the cells having the selectable mutation will be found to also contain a mutation in the second target gene that was introduced by the second MDON. More frequently at least about 10% of the cells having the selectable mutation will be found to also contain a mutation in the second target gene.

One use of this embodiment of the invention is the investigation of the function of a gene-of-interest. A mixture is provided of a MDON that causes a selectable mutation and a MDON that causes a mutation that would be expected to "knock-out" the gene-of-interest, e.g., the insertion of a stop codon or a frameshift mutation. Cells in which one or more copies of the gene-of-interest have been knocked out can be recovered from the population having the selectable mutation. Such cells can be regenerated into a plant so that the function of the gene-of-interest can be determined.

A selectable trait can be caused by any mutation that causes a phenotypic change that can produce a selective growth advantage under the appropriate selective conditions or a phenotypic change that can be readily observed, such as change in color of the plant cells growing in a callus. The selectable trait can itself be a desirable traits, e.g., herbicide resistance, or the selectable trait can be used merely to facilitate the isolation of plants having a non-selectable trait that was introduced by transmutation. A desired nonselectable trait can be introduced into a cell by using a mixture of the MDON that causes the desired mutation and the MDON that causes the selectable mutation, followed by culture under the selecting conditions. Selection according to this scheme has the advantage of ensuring that each selected cell not only received the mixture of MDONs, but also that the cell which received the mixture was then susceptible to transmutation by a MDON.

A mutation that causes a lethal phenotypic change under the appropriate conditions, termed a negatively selectable mutation, can also be used in the present invention. Such mutations cause negatively selectable traits. Negatively selectable

traits can be selected by making replica plates of the transmutated cells, selecting one of the replicas and recovering the transmutated cell having the desired property from the non-selected replica.

4.6 Specific Genes That Can Be Transmutated to Create Selectable Traits

In one embodiment of the invention a MDON is used to introduce a mutation into an Acetolactate synthase (ALS) gene, which is also termed the aceto-hydroxy amino acid synthase (AHAS) gene. Sulfonyleurea herbicides and imidazoline herbicides are inhibitors of the wild type ALS enzymes. Dominant mutations that render plants resistant to the actions of sulfonyleureas and imidazolines have been described. See U.S. Patent Nos. 5,013,659 and 5,378,824 (Bedbrook) and Rajasekaran K., et al., 1996, Mol. Breeding 2, 307-319 (Rajasekaran). Bedbrook at Table 2 describes several mutations (hereafter, a "Bedbrook Mutation") that were found to render yeast ALS enzymes resistant to sulfonyleurea herbicides. Bedbrook states that each of the Bedbrook mutations makes a plant resistant to sulfonyleurea and imidazoline herbicides when introduced into a plant ALS gene. It is understood that in most plants the gene encoding ALS has been duplicated. A mutation can be introduced into any allele of either ALS gene.

Three of the Bedbrook mutations were, in fact, shown to confer herbicide resistance in a plant, namely the substitutions Pro→Ala¹⁹⁷, Ala→Asp²⁰⁵ and Trp→Leu⁵⁹¹. Rajasekaran reports that mutations Trp→Ser⁵⁹¹ caused resistance to sulfonyleurea and imidazoline and that Ser→Asn⁶⁶⁰ caused resistance to imidazoline herbicides. The results of Rajasekaran are reported herein using the sequence numbering of Bedbrook. Those skilled in the art will understand that the ALS genes of different plants are of unequal lengths. For clarity, a numbering system is used in which homologous positions are designated by the same position number in each species. Thus, the designated position of a mutation is determined by the sequence that surrounds it. For example, the mutation Trp→Ser⁵⁹¹ of Rajasekaran is at residue 563 of the cotton ALS gene but is designated as position 591 of Bedbrook because the mutated Trp is surrounded by the sequence that surrounds Trp⁵⁹¹ in Table 2 of Bedbrook. According to the invention any substitution for the naturally occurring amino acid at position 660 or one of the positions listed in Table 2 of Bedbrook, which is hereby incorporated by

reference, can be used to make a selectable mutation in the ALS gene of a plant.

In a further embodiment of the invention the selectable mutation can be a mutation in the chloroplast gene *psbA* that encodes the D1 subunit of photosystem II, see Hirschberg, J., et al., 1984, *Z. Naturforsch.* **39**, 412-420 and Ohad, N., & Hirschberg, J., *The Plant Cell* **4**, 273-282. Hirschberg et al. reports that the mutation Ser-Gly²⁶⁴ results in resistance to triazine herbicides, e.g., 2-Cl-4-ethylamino-6-isopropylamino-s-triazine (Atrazine). Other mutations in the *psbA* gene that cause Atrazine herbicide resistance are described in Erickson J.M., et al., 1989, *Plant Cell* **1**, 361-371, (hereafter an "Erickson mutation"), which is hereby incorporated by reference. The use of the selectable trait caused by an Erickson mutation is preferred when it is desired to introduce a second new trait into a chloroplast.

The scientific literature contains further reports of other mutations that produce selectable traits. Ghislain M., et al., 1995, *The Plant Journal* **8**, 733-743, describes a Asn-Ile¹⁰⁴ mutation in the *Nicotiana sylvestris* dihydrodipicolinate synthase (DHDPs, EC 4.2.1.52) gene that results in resistance to S-(2-aminoethyl)L-cysteine. Mourad, G., & King, J., 1995, *Plant Physiology* **109**, 43-52 describes a mutation in the threonine dehydratase of *Arabidopsis thaliana* that results in resistance to L-O-methylthreonine. Nelson, J.A.E., et al., 1994, *Mol. Cell. Biol.* **14**, 4011-4019 describes the substitution of the C-terminal Leu of the S14/rp59 ribosomal protein by Pro, which causes resistance to the translational inhibitors cryptoluerine and emetine. In further embodiments of the invention, each of the foregoing mutations can be used to create a selectable trait. Each of Ghislain, Mourad and Nelson are hereby incorporated by reference.

4.7 Genes That Can Be Mutated to Create Desirable Non-selectable Traits

Example 1

MALE STERILITY

Certain commercially grown plants are routinely grown from hybrid seed including corn (maize, *Zea mays*), tomatoes and most other vegetables. The production of hybrid seed requires that plants of one purebred line be pollinated only by pollen from another purebred line, i.e., that there be no self pollination. The removal of the pollen-producing organs from the purebred parental plants is a

laborious and expensive process. Therefore, a mutation that induces male-sterility i.e., suppresses pollen production or function, would obviate the need for such process.

Several genes have been identified that are necessary for the maturation or function of pollen but are not essential for other processes of the plant. Chalcone synthase (*chs*) is the key enzyme in the synthesis of flavonoids, which are pigments found in flowers and pollen. Inhibition of *chs* by the introduction of a *chs* antisense expressing gene in the petunia results in male sterility of the plant. Van der Meer, I.M., et al., 1992, *The Plant Cell* **4**, 253-262. There is a family of *chs* genes in most plants. See, e.g., Koes, R.E., et al., 1989, *Plant Mol. Biol.* **12**, 213-226. Likewise disruption of the chalcone synthase gene in maize by insertion of a transposable element results in male sterility. Coe, E.H., *J. Hered.* **72**, 318-320. The structure of maize chalcone synthase and a duplicate gene, *whp*, is given in Franken, P., et al., 1991, *EMBO J.* **10**, 2605-2612. Typically in plants each member of a multigene family is expressed only in a limited range of tissues. Accordingly, the present embodiment of the invention requires that in species having multiple copies of chalcone synthase genes, the particular *chs* gene or genes expressed in the anthers be identified and interrupted by introduction of a frameshift, and one or more in-frame termination codons or by interruption of the promoter.

A second gene that has been identified as essential for the production of pollen is termed *Lat52* in tomato. Muschietti, J., et al., 1994, *The Plant Journal* **6**, 321-338. *LAT52* is a secreted glycoprotein that is related to a trypsin inhibitor. Homologs of *Lat52* have been identified in maize (termed *Zm13*, Hanson D.D., et al., 1989 *Plant Cell* **1**, 173-179; Twell D., et al., 1989, *Mol. Gen. Genet.* **217**, 240-245), rice (termed *Ps1*, Zou J., et al., 1994 *Am. J. Bot.* **81**, 552-561 and olive (termed *Ole e I*, Villalba, M., et al., 1993, *Eur. J. Biochem.* **276**, 863-869). Accordingly, the present embodiment of the invention provides for a method of obtaining male sterility by the interruption of the *Lat52/Zm13* gene or its homologs by the introduction of a frameshift, one or more in-frame termination codons or by interruption of the promoter.

A third gene that has been identified as essential for the production of pollen is the gene that encodes phenylalanine ammonium lyase (PAL, EC 4.3.1.5). PAL is an essential enzyme in the production of both phenylpropanoids and flavonoids.

Because phenylpropanoids are a precursor to lignins, which can be an essential for the resistance to disease in the preferred embodiment a PAL isozyme that is expressed only in the anther is identified and interrupted to obtain male sterility.

Example 2 ALTERATION OF CARBOHYDRATE METABOLISM OF TUBERS

Once harvested, potato tubers are subject to disease, shrinkage and sprouting during storage. To avoid these losses the storage temperature is reduced to 35-40° F. However, at reduced temperatures, the starch in the tubers undergoes conversion to sugar, termed "cold sweetening", which reduces the commercial and nutritional value of the tuber. Two enzymes are critical for the cold sweetening process: acid invertase and UDP-glucose pyrophosphorylase. Zrenner, R., et al., 1996, *Planta* **198**, 246-252 and Spychalla, J.P., et al., 1994, *J. Plant Physiol.* **144**, 444-453, respectively. The sequence of potato acid invertase is found in EMBL database Accession No. X70368 (SEQ ID NO. 1) and the sequence of the potato UDP Glucose pyrophosphorylase is reported by Katsube, T. et al., 1991, *Biochem.* **30**, 8546-8551. Accordingly, the present embodiment of the invention provides for a method of preventing cold sweetening by the interruption of the acid invertase or the UDP glucose phosphorylase gene by introduction of a frameshift, one or more in-frame termination codons or by interruption of the promoter.

Example 3 REDUCTION IN POST HARVEST BROWNING DUE TO PPO

Polyphenol oxidase (PPO) is the major cause of enzymatic browning in higher plants. PPO catalyzes the conversion of monophenols to o-diphenols and of o-dihydroxyphenols to o-quinones. The quinone products then polymerize and react with amino acid groups in the cellular proteins, which results in discoloration. The problem of PPO induced browning is routinely addressed by the addition of sulfites to the foods, which has been found to be associated with some possible health risk and consumer aversion. PPO normally functions in the defense of the plant to pathogens or insect pests and, hence, is not essential to the viability of the plant. Accordingly, the present embodiment of the invention provides for a method of preventing enzymatic browning by the interruption of the PPO gene by introduction of a frameshift, one or more in-frame termination codons or by interruption of the promoter

in apple, grape, avocado, pear and banana.

The number of PPO genes in the genome of a plant is variable; in tomatoes and potatoes PPO forms a multigene family. Newman, S.M., et al., 1993, *Plant Mol. Biol.* **21**, 1035-1051, Hunt M.D., et al., 1993, *Plant Mol. Biol.* **21**, 59-68; Thygesen, P.W., et al., 1995, *Plant Physiol.* **109**, 525-531. The grape contains only a single PPO gene. Dry, I.B., et al., 1994, *Plant Mol. Biol.*, **26**, 495-502. When the plant species of interest contains multiple copies of PPO genes it is essential that the PPO gene that is normally expressed in the commercial product be interrupted. For example, only one PPO gene is expressed in potatoes of harvestable size, which gene is termed POT32 and its sequence is deposited in GENBANK accession No. U22921 (SEQ ID NO. 2), which sequence is incorporated by reference. The other potato PPO isozymes have been sequenced and the sequences deposited so that one skilled in the art can design a MDON that specifically inactivates POT32.

Example 4 REDUCTION OF LIGNIN IN FORAGE CROPS AND WOOD PULP

Lignin is a complex heterogeneous aromatic polymer, which waterproofs higher plants and strengthens their cell walls. Lignin arises from the random polymerization of free radicals of phenylpropanoid monolignins. Lignins pose a serious problem for the paper industry because their removal from wood pulp involves both monetary and environmental costs. Similarly, the lignin content of forage crops limits their digestibility by ruminants. Indeed, naturally occurring mutations, termed "brown mid-rib" in sorghum, Porter, KS, et al., 1978, *Crop Science* **18**, 205-218, and maize, Lechtenberg, V.L., et al., 1972, *Agron. J.* **64**, 657-660, have been identified as having reduced lignin content and tested as feed for cattle.

The brown mid-rib mutation in maize involves the O-methyl transferase gene. Vignol, F., et al., 1995, *Plant Cell* **7**, 407-416. The O-methyltransferase genes of a number of plant species have been cloned: Burgos, R.C., et al., 1991, *Plant Mol. Biol.* **17**, 1203-1215 (aspen); Gowri, G., et al., 1991, *Plant Physiol.* **97**, 7-14 (alfalfa, *Medicago sativa*) and Jaeck, E., et al., 1992, *Mol. Plant-Microbe Interact.* **4**, 294-300 (tobacco) (SEQ ID No. 3 and SEQ ID No. 4). Thus, one aspect of the present embodiment is the interruption of the O-methyltransferase gene to reproduce a brown mid-rib phenotype in any cultivar of maize or sorghum and in other species of forage

crops and in plants intended for the manufacture of wood pulp.

A second gene that is involved in lignin production is the cinnamyl alcohol dehydrogenase (CAD) gene, which has been cloned in tobacco. Knight, M.E., 1992, Plant Mol. Biol. 19, 793-801 (SEQ ID No. 5 and SEQ ID No. 6). Transgenic tobacco plants making a CAD antisense transcript have reduced levels of CAD and also make a lignin that is more readily extractable, apparently due to an increase in the ratio of syringyl to guaiacyl monomers and to the increased incorporation of aldehyde monomers relative to alcohol residues. Halpin, C., et al., 1994, The Plant Journal 6, 339-350. Accordingly, an embodiment of the invention is the interruption of the CAD gene of forage crops such as alfalfa, maize, sorghum and soybean and of paper pulp trees such as short-leaf pine (*Pinus echinata*) long-leaf pine (*Pinus palustris*) slash pine (*Pinus elliottii*), loblolly pine (*Pinus taeda*), yellow-poplar (*Liriodendron tulipifera*) and cotton wood (*Populus sp.*) by introduction of a frameshift, one or more in-frame termination codons or by interruption of the promoter.

Example 5 THE REDUCTION IN UNSATURATED AND POLYUNSATURATED LIPIDS IN OIL SEEDS

The presence of unsaturated fatty acids, e.g., oleic acid, and polyunsaturated fatty acids, e.g., linoleic and linolenic acids, in vegetable oil from oil seeds such as rape, peanut, sunflower and soybean causes the oils to oxidize, on prolonged storage and at high temperatures. Consequently, vegetable oil is frequently hydrogenated. However, chemical hydrogenation causes transhydrogenation, which produces non-naturally occurring stereo-isomers, which are believed to be a health risk.

Fatty acid synthesis proceeds by the synthesis of the saturated fatty acid on an acyl carrier protein (ACP) followed by the action of desaturases that remove the hydrogen pairs. Consequently, it would be desirable to inhibit the activity of these desaturase enzymes in oil seeds.

The first enzyme in the synthesis of oleic acid is stearyl-ACP desaturase (EC 1.14.99.6). The stearyl-ACP desaturases from safflower and castor bean have been cloned and sequenced. Thompson, G.A., et al., 1991, Proc. Natl. Acad. Sci. 88, 2578-2582 (SEQ ID No. 7); Shanklin, J., & Somerville, C., 1991, Proc. Natl. Acad. Sci. 88, 2510-2514 (SEQ ID No. 8); Knutzon, D.S., et al., 1991, Plant Physiology 96, 344-

345. Accordingly, one embodiment of the present invention is the interruption of the stearoyl-ACP desaturase gene of oil seed crops such as soybean, safflower, sunflower, soy, maize and rape by introduction of a frameshift, one or more in-frame termination codons or by interruption of the promoter.

A second enzyme that can be interrupted according to the present invention is ω -3 fatty acid desaturase (ω -3 FAD) the enzyme that converts linoleic acid, a diene, to linolenic acid, a triene. There are two ω -3 FAD isozymes in *Arabidopsis thaliana* and, those skilled in the art expect, in most other plants. One isozyme is specific for plastids and is the relevant isozyme for the synthesis of the storage oils of seeds. The other is microsome specific. The cloning of the *Arabidopsis thaliana* plastid ω -3 FAD is reported by Iba., K. et al., 1993, J. Biol. Chem. 268, 24099-24105 (SEQ ID No. 9). Accordingly an embodiment of the invention is the interruption of the plastid ω -3 FAD gene of oil seed crops such as soybean, safflower, sunflower, soy, maize and rape by introduction of a frameshift, an in-frame termination codon or by interruption of the promoter.

Example 6 INACTIVATION OF S ALLELES TO PERMIT INBRED LINES

Certain plant species have developed a mechanism to prevent self-fertilization. In these species, e.g., wheat and rice, there is a locus, termed S, which has multiple alleles. A plant that expresses an S allele cannot be fertilized by pollen expressing the same S allele. Lee, H-K., et al., 1994, Nature 367, 560-563; Murfett, J., et al., 1994, Nature 367, 563. The product of the S locus is an RNase. McClure, B.A., et al., 1989, Nature 342, 955-957. The product of the S locus is not essential for the plant. Accordingly, an embodiment of the invention is the interruption of genes of the S locus to permit the inbreeding of the plant by introduction of a frameshift, one or more in-frame termination codons or by interruption of the promoter.

Example 7 ETHYLENE INSENSITIVITY

Ethylene is a gaseous plant hormone that is involved in plant growth and development. An unwanted aspect of ethylene's action is the over-ripening of fruit, vegetables and the wilting of flowers that results in rotting and loss. The ethylene

receptor of *Arabidopsis thaliana* has been cloned and is termed ETR-1. Chang, C., et al., 1993, Science 262, 539-544 (SEQ ID No. 10). A mutant, Cys-Tyr⁶⁵, results in a dominant insensitivity to ethylene. Transgenic tomato plants expressing the *Arabidopsis thaliana* mutant ETR-1 also showed an insensitivity to ethylene, indicating that the Cys-Tyr⁶⁵ mutation would be a dominant suppressor of ethylene action in most plant species. Accordingly one aspect of the present embodiment of the invention is the insertion of the Cys-Tyr⁶⁵ mutation into the ETR-1 gene so as to extend the life span of the mutated fruit vegetable or flower.

In a further aspect of the present embodiment, the preservation of the fruit or flower can be achieved by interrupting one of the genes that encode the enzymes for ethylene synthesis: namely 1-aminocyclopropane-1-carboxylic acid synthase (ACC synthase) and ACC oxidase. For this embodiment of the invention the amount of ethylene synthesis can be eliminated entirely, so that ripening is produced by exogenous ethylene or some amount of ethylene production can be retained so that the fruit ripens spontaneously, but a has a prolonged storage life. Accordingly, it is anticipated that the interruption of one allele of either the ACC synthase or the ACC oxidase gene can result in an useful reduction in the level of ethylene synthesis. Alternatively, the invention provides for the interruption of one allele along with the introduction of a mutation that results in a partial loss of activity in the uninterrupted allele.

The sequences of the *Arabidopsis thaliana* ACC synthase and ACC oxidase genes are reported in Abel., S., et al., 1995, J. Biol. Chem. 270, 19093-19099 (SEQ ID No. 12) and Gomez-Lim, M.A., et al., 1993, Gene 134, 217-221 (SEQ ID No. 11), respectively, which are incorporated by reference in their entirety.

Example 8 REVERSION OF KANAMYCIN RESISTANCE

Recombinant DNA technology in plants allows for the introduction of genes from one species of plant and bacterial genes into a second species of plant. For example, Kinney, A.J., 1996, Nature Biotech. 14, 946, describes the introduction of a bay laural ACP-thioesterase gene into the rape seed to obtain a vegetable oil rich in lauric acid. Such transgenic plants are normally constructed using an antibiotic resistance gene, e.g., kanamycin resistance, which is coinserted into the transgenic

plant as a selectable trait. The resultant transgenic plant continues to express the antibiotic resistance gene, which could result in large amounts of the resistance product and the gene entering the food supply and/or the environment, which introduction may represent an environmental or health risk. An embodiment of the invention obviates the risk by providing for the interruption of the kanamycin gene by introduction of a frameshift, one or more in-frame termination codons or by interruption of the promoter.

Example 9 **MODIFICATION OF STORAGE PROTEIN AMINO ACID CONTENT**

Seeds and tubers contain a family of major storage proteins, e.g., patatins in potato and zeins in maize. The amino acid composition of such storage proteins is often poorly suited to the needs of the human and animals that depend on these crops, e.g., corn is deficient in lysine and methionine and potato is deficient in methionine. Accordingly, one embodiment of the invention is the mutation of a storage protein of a food crop to increase the amount of low abundance amino acids. Patatins are encoded by a multigene family which are characterized in Mignery, G.A., et al., 1988, *Gene* 62, 27-44, and the structure of zeins is reported by Marks, M.D., et al., 1985, *J. Biol. Chem.* 260, 16451-16459, both of which are hereby incorporated by reference. Alternatively, the anticodon of a methionine or lysine specific tRNA can be mutated to that of a more common amino acid.

Example 10 **THE USE OF MDON TO DETERMINE THE FUNCTION OF A GENE**

The presently available techniques for the cloning and sequencing of tissue specific cDNAs allow those skilled in the art to obtain readily the sequences of many genes. There is a relative paucity of techniques for determining the function of these genes. In one embodiment of the invention, MDON are designed to introduce frameshift or stop codons into the gene encoding a cDNA of unknown function. This allows for the specific interruption of the gene. Plants having such specific "knock-outs" can be grown and the effects of the knock-out can be observed in order to investigate the function of the unknown gene.

4.8 Fertile Plants of the Invention

The invention encompasses a fertile plant having an isolated selectable point mutation, which isolated selectable mutation is not a rare polymorphism, i.e., would not be found in population of about 10,000 individuals. As used herein a point mutation is mutation that is a substitution of not more than six contiguous nucleotides, preferably not more than three and more preferably one nucleotide or a deletion or insertion from one to five nucleotides and preferably of one or two nucleotides. As used herein an isolated mutation is a mutation which is not closely linked genetically to any other mutation, wherein it is understood that mutations that are greater than 100 Kb and preferably greater than 40 Kb and more preferably greater than 23 Kb are not closely linked.

BIOLISTIC WORKING EXAMPLES

In the following working examples the media and protocols found in Gelvin, S.B., et al., (eds) 1991, PLANT MOLECULAR BIOLOGY MANUAL (Kluwer Acad. Pub.) were followed. Gold particles were coated with MDON according the following protocol. The microprojectiles are first prepared for coating, then immediately coated with the chimeraplast. To prepare the microprojectiles, suspend 60 mg of gold particles in 1 ml of 100% ethanol (see Note 4). Sonicate the suspension for three, 30 s bursts to disperse the particles. Centrifuge at 12,000 xg for 30 s, discard supernatant. Add 1 ml of 100% ethanol, vortex for 15 s, centrifuge at 12,000 xg for 5 min, then discard the supernatant. A 25 μ l suspension of washed gold particles (1.0 μ m diameter, 60 mg/ml) in H₂O are slowly vortexed, to which 40 μ l MDON (50 μ g/ml), 75 μ l of 2.5 M CaCl₂, 75 μ l 0.1M spermidine are sequentially added. All solutions are ice cold. The completed mixture is vortexed for a further 10 min and the particles are allowed to settle at room temperature for a further 10 min. The pellet is washed in 100% EtOH and resuspended in 50 μ l. of absolute ethanol. Biolistic delivery is performed using a Biorad Biolistic gun with the following settings: tank pressure 1100 psi, rupture disks x2 breaking at 900 psi, particle suspension volume 5 μ l.

NT-1 (TOBACCO), A DICOT CELL SUSPENSION: Lawns of NT-1 of approximately 5 cm diameter, containing 5 million cells, were grown for 3 days on standard media at

28°C. Gold particles were coated with ALS-1 or ALS-2 and were shot as above. The cells were cultured a further 2.5 days, suspended and transferred to solid medium supplemented with 15-50 ppb chlorosulfuron (GLEAM™). Resistant colonies emerged after 7-14 days.

The sequences of the MDON used are as follows: (The nucleotides not homologous with the target gene are underlined and bold. Lower case letters denote 2'-O-methyl ribonucleotides.)

ALS-1

```

TGCGCG-guccaguucaCGTTGcauccaacuaT
T                                     T
T                                     T (SEQ ID No. 13)
TCGCGC CAGGTCAAGTGCAACGTAGGATGATT
      3' 5'

```

ALS-2

```

TGCGCG-guccaguucaCGATGcauccaacuaT
T                                     T
T                                     T (SEQ ID No. 14)
TCGCGC CAGGTCAAGTGCTACGTAGGATGATT
      3' 5'

```

ALS-1 and ALS-2 have single base mismatches with the ALS gene at the second nucleotide of the Pro¹⁹⁷ (CCA) codon: ALS-1 is CAA and ALS-2 is CTA. Following PCR amplification and sequencing of the gene of the ALS-1 and ALS-2 transmutated, resistant cell lines, a mutation was in the targeted codon which was found to be Thr (ACA) and Ser (TCA), respectively. The observed mutation was shifted one nucleotide 5' of the location that would have been expected based on the action of MDON in mammalian cells on the coding strand and one nucleotide 3' of the expected location on the non-coding strand. A total of 3 ALS-1 and 5 ALS-2 transmutants having these mutations were identified. No resistant calli were obtained from ALS-1 DNA treated cells.

For selection of chlorosulfuron resistant cells, cells were transferred from each bombarded plate to 15 ml containing 5 ml of liquid CSM 2 d after bombardment. The tubes were inverted several times to disperse cell clumps. The cells were then transferred to solidified CSM medium containing 15 ppb chlorosulfuron (Dupont, Wilmington, DE). After approximately 3 - 5 wk, actively growing cells (raised, light

colored colonies) are selected and transferred to solidified CSM containing 50 ppb chlorsulfuron. Three to four weeks later, actively growing cells are selected, then transferred to solidified CSM containing 200 ppb chlorsulfuron. Cells that survive this treatment are then analyzed.

MEDIA

1. NT-1 cell suspension medium (CSM): Murashige and Skoog salts (Gibco BRL, Grand Island, NY), 500 mg/l MES, 1 mg/l thiamine, 100 mg/l myoinositol, 180 mg/l KH_2PO_4 , 2.21 mg/L 2,4-dichlorophenoxyacetic acid (2,4-D), 30g/L sucrose. Adjust pH to 5.7 with 1M KOH or HCl and autoclave. For solidified medium add 8g/l Agar-agar (Sigma, St. Louis, MO) prior to autoclaving.
2. Plating out medium (POM): 80% (v/v) CSM, 0.3M mannitol, 20% (v/v) supernatant from the initial centrifugation of the NT-1 cell suspension prior to protoplast isolation.

TOBACCO LEAF, A DICOT: *Nicotiana tabacum* v. *Samsun* leaf disks were co-transformed by *Agrobacterium tumefaciens* LBA 4404 harboring bin 19-derived plasmids containing a nptII expression cassette containing two genes: a gene for kanamycin resistance and one of two mutants of a gene encoding a Green Fluorescence Protein (GFP, Chui, W., 1996, Current Biol. 6, 325-330). Neither mutant GFP gene produces a GFP product. The mutants contain either a G→T substitution in the sixth codon resulting in a stop codon or a deletion of one nucleotide at the same position, which are termed, respectively, G-stop and G-Δ. After culture on selective MS 104 medium, leaves were recovered and the presence of a GFP gene confirmed by northern blot.

Sequence of first eight codons of GFP:

GFP	ATG GTG AGC AAG GGC GAG GAG CTG	(SEQ ID No. 15)
G-stop	—————T—————	(SEQ ID No. 16)
G-Δ	—————AGG AGC TGT	(SEQ ID No. 17)

The sequences of the MDON used were as follows: (The nucleotides not homologous with G-stop are underlined and bold. Lower case letters denote 2'-O-methyl ribonucleotides.)

GFP-1

TGCGCG-cacucguuccCGCTCcucgacaaguT
T T (SEQ ID No. 18)
T T
TCGCGC GTGAGCAAGGGCGAGGAGCTGTTTCAT
3' 5'

GFP-2

TGCGCG-acucguucccGAGCCucgacaagugT
T T (SEQ ID NO. 19)
T T
TCGCGC TGAGCAAGGGCTCGGAGCTGTTCACT
3' 5'

Leaf disks of the G-stop and G-Δ transgenic plants were incubated on MS 104 selective media and G-1 or G-1 introduced biolistically by two successive deliveries as above. Approximately 10 days after the introduction of the MDON, calli exhibiting GFP-like fluorescence were seen in the G-1 and G-2 treated cultures of both the G-stop and G-Δ leaf disks. Larger and more rapidly growing callusing pieces were subdivided by scalpel to obtain green fluorescent cell-enriched calli. The fluorescent phenotype remained stable for the total period of observation, about 30 days. The presence of green fluorescent cells in the G-1 treated G-stop culture indicates that G-1 does not cause mutations exclusively one base 5' of the mismatched nucleotide.

Green fluorescence was observed using a standard FITC filter set using an IMT-2 Olympus microscope.

ELECTROPORATION WORKING EXAMPLE

CONVERSION OF GFP IN TOBACCO MESOPHYLL PROTOPLASTS

Plant Material

1. Tobacco plant transformant (Delta6) harboring a deletion mutant of GFP.
2. Leaves were harvested from 5 to 6-week-old *in vitro*-grown plantlets

Protoplast Isolation

1. Basically followed the procedure of Gallois, et al., 1996, Electroporation of tobacco leaf protoplasts using plasmid DNA or total genomic DNA. Methods in Molecular Biology, Vol. 55: Plant Cell Electroporation and Electrofusion Protocols Edited by: J. A.

Nickoloff Humana Press Inc., Totowa, NJ. pp.89 - 107.

2. Enzyme solution: 1.2 % cellulase R-10 "Onozuka" (Karlson, Santa Rosa, CA), 0.8% macerozyme R-10 (Karlson, Santa Rosa, CA), 90 g/l mannitol, 10 mM MES, filter sterilize, store in 10 ml aliquots at -20°C.
3. Leaves were cut from the mid-vein out every 1 - 2 mm. They were then placed abaxial side down in contact with 10 ml of enzyme solution in a 100 x 20 mm petri plate. A total of 1 g of leaves was placed in each plate.
4. The plates were incubated at 25°C in the dark for 16 hr.
5. The digested leaf material was pipetted and sieved through a 100 μ m nylon screen cloth (Small Parts, Inc., Miami Lakes, FL). The filtrate was then transferred to a centrifuge tube, and centrifuged at 1000 rpm for 10 min. All centrifugations for this protocol were done at these conditions.
6. The protoplasts collected in a band at the top. The band of protoplasts was then transferred to a clean centrifuge to which 10 ml of a washing solution (0.4 M sucrose and 80 mM KCl) was added. The protoplasts were gently resuspended, then centrifuged.
7. Repeated step 6 twice.
8. After the last wash, the protoplast density was determined by dispensing a small aliquot onto a hemocytometer. Resuspend the protoplasts to a density of 1×10^6 protoplasts/ml in electroporation buffer (80 mM KCl, 4 mM CaCl_2 , 2mM potassium phosphate, pH 7.2, 8% mannitol, autoclave. The protoplasts were allowed to incubate at 8°C for 2 hr.
9. After 2 hr, 0.3 ml (3×10^5 protoplasts) were transferred to each 0.4 cm cuvette, then placed on ice. GFP-2 (0.6 - 4 μ g/mL) was added to each cuvette except for an unelectroporated control. The protoplasts were electroporated (250V, capacitance 250 μ F, and time constant 10 - 14 ms).
10. The protoplasts were allowed to recover for 10 min on ice, then transferred to petri

plates (100 x 20 mm). After 35 min, 10 ml of POM, see above, was added to each plate. The plates were transferred to the dark at 25°C for 24 hr, then transferred to the light.

11. The protoplast cultures were then maintained according to *Gallois supra*.

Fluorescence Microscopy

1. Under UV light, we observed 8 GFP converted protoplasts out of 3×10^5 protoplasts.

We Claim:

1. A method of making a localized mutation in a target gene in a plant cell comprising the steps of:
 - a. adhering to a particle a recombinagenic oligonucleobase, which contains a first homologous region which has a sequence identical to the sequence of at least 6 base pairs of a first fragment of the target gene and a second homologous region which has a sequence identical to the sequence of at least 6 base pairs of a second fragment of the target gene, and an intervening region which contains at least 1 nucleobase heterologous to the target gene, which intervening region connects the first homologous region and the second homologous region;
 - b. introducing the particle into a cell of a population of plant cells;
 - c. identifying a cell of the population cell having a mutation located between the first and second fragments of the target gene.
2. The method of claim 1, wherein the recombinagenic oligonucleobase is a MDON and each of the homologous regions contains an RNA segment of at least 6 RNA-type nucleotides.
3. The method of claim 2, wherein the intervening region is at least 3 nucleotides in length.
4. The method of claim 2, which further comprises the step of culturing the identified cell so that a plant is generated.
5. The method of claim 2, wherein the first RNA segment contains at least 8 contiguous 2'-Substituted Ribonucleotides.
6. The method of claim 5 wherein the second RNA segment contains at least 8 contiguous 2'-Substituted Ribonucleotides.
7. The method of claim 2, wherein the sequence of the mutated target gene is homologous with the sequence of the MDON.
8. The method of claim 2, wherein the adhering step is performed in a solution

comprising 1.1-1.4 M NaCl and 18-22 μ M spermidine and at least 14 μ g/ml MDON.

9. The method of claim 2, wherein the target gene is a first ALS gene, a second ALS gene, a psbA gene, a threonine dehydratase gene, a dihydrodipicolinate synthase gene, or an S14/rp59 gene
10. The method of claim 9, wherein the plant cell is a maize, wheat, rice or lettuce cell.
11. The method of claim 9, wherein the plant cell is a potato, tomato, canola, soybean or cotton cell.
12. The method of claim 2, wherein the target gene selected from the group consisting of the genes encoding acid invertase, UDP-glucose pyrophosphorylase, polyphenol oxidase, O-methyl transferase, cinnamyl alcohol dehydrogenase, *etr-1* or a homolog thereof, ACC synthase and ACC oxidase.
13. The method of claim 12, where the plant cell is from a maize, wheat, rice or lettuce plant.
14. The method of claim 12, where the plant cell is from a potato, tomato, canola, soybean or cotton plant.
15. The method of claim 2, which further comprises making seeds from the plant or from progeny of the plant.
16. A method of making a localized mutation in a target gene in a plant cell having a cell wall comprising the steps of:
 - a. perforating the cell walls of a population of plant cells;
 - b. introducing a recombinagenic oligonucleobase, which contains a first homologous region which has a sequence identical to the sequence of at least 6 base pairs of a first fragment of the target gene and a second homologous region which has a sequence identical to the sequence of at least 6 base pairs of a second fragment of the target gene, and an intervening region which contains at least 1 nucleobase heterologous to the target gene, which intervening region connects the first homologous region

and the second homologous region;

- c. identifying a cell of the population having a mutation located between the first and second fragments of the target gene.
17. The method of claim 16, wherein the recombinagenic oligonucleobase is a MDON and each of the homologous regions contains an RNA segment of at least 6 RNA-Type nucleotides.
 18. The method of claim 17, which further comprises the step of culturing the identified cell so that a plant is generated.
 19. The method of claim 17, wherein the sequence of the target gene between the first and the second fragments differs from the sequence of the intervening region of the MDON at a mismatched nucleotide and the mutation of the target gene is located adjacent to the mismatched nucleotide.
 20. The method of claim 17, wherein the sequence of the target gene between the first and the second fragments differs from the sequence of the mutator segment of the MDON at a mismatched nucleotide and the mutation of the target gene is located at the mismatched nucleotide.
 21. The method of claim 17, wherein the target gene is a first ALS gene, a second ALS gene, a psbA gene, a threonine dehydratase gene, a dihydrodipicolinate synthase gene, or an S14/rp59 gene
 22. The method of claim 21, wherein the plant cell is a maize, wheat, rice or lettuce cell.
 23. The method of claim 21, wherein the plant cell is a potato, tomato, canola, soybean or cotton cell.
 24. The method of claim 17, wherein the target gene is selected from the group consisting of the genes encoding acid invertase, UDP-glucose pyrophosphorylase, polyphenol oxidase, O-methyl transferase, cinnamyl alcohol dehydrogenase, *etr-1* or a homolog thereof, ACC synthase and ACC oxidase.
 25. The method of claim 24, where the target gene is a gene from a maize, wheat, rice or lettuce plant.

26. The method of claim 24, where the target gene is a gene from a potato, tomato, canola, soybean or cotton plant.
27. The method of claim 17, which further comprises making seeds from the plant or from progeny of the plant.
28. A method of making a localized mutation in a target gene of a plastid of a plant cell which comprises the steps of:
 - a. Introducing a recombinagenic oligonucleobase, which contains a first homologous region which has a sequence identical to the sequence of at least 6 base pairs of a first fragment of the target gene and a second homologous region which has a sequence identical to the sequence of at least 6 base pairs of a second fragment of the target gene, and an intervening region which contains at least 1 nucleobase heterologous to the target gene, which intervening region connects the first homologous region and the second homologous region;
 - b. Identifying a cell having a mutation in the region between the first and second fragments of the target gene.
29. The method of claim 28, wherein the recombinagenic oligonucleobase is a MDON and each of the homologous regions contains an RNA segment of at least 6 RNA-Type nucleotides.
30. The method of claim 29, which further comprises culturing the identified cell so that a plant is generated.
31. A method of making a localized, non-selectable mutation in a target gene in a plant cell comprising the steps of:
 - a. introducing into the cells of a population of cells a mixture of a first recombinagenic oligonucleobase and a second recombinagenic oligonucleobase wherein:
 - i. the first recombinagenic oligonucleobase contains a first homologous region which has a sequence identical to the sequence of at least 6 base pairs of a first fragment of a first target gene and a second homologous

region which has a sequence identical to the sequence of at least 6 base pairs of a second fragment of the first target gene, and an intervening region which contains at least 1 nucleobase heterologous to the target gene, which intervening region connects the first homologous region and the second homologous region, and

- ii. the second recombinagenic oligonucleobase contains a first homologous region which has a sequence identical to the sequence of at least 6 base pairs of a first fragment of a second target gene and a second homologous region which has a sequence identical to the sequence of at least 6 base pairs of a second fragment of the second target gene, and an intervening region which contains at least 1 nucleobase heterologous to the target gene, which intervening region connects the first homologous region and the second homologous region;
- b. selecting cells from the population having a selectable mutation located between the first and the second fragments of the first target gene from the population; and
- c. identifying a selected cell having a non-selectable mutation located between the first fragment and the second fragment of the second target cell.

- 32. The method of claim 31, wherein the each recombinagenic oligonucleobase is a MDON and each of the homologous regions contains an RNA segment of at least 6 RNA-Type nucleotides.
- 33. The method of claim 32, wherein the first target gene is a first ALS gene, a second ALS gene, a psbA gene, a threonine dehydratase gene, a dihydrodipicolinate synthase gene, or an S14/rp59 gene.
- 34. The method of claim 33, wherein the plant cell is a maize, wheat, rice or lettuce cell.
- 35. The method of claim 33, wherein the plant cell is a potato, tomato, canola, soybean or cotton cell.

36. The method of claim 32, wherein the second target gene is selected from the group consisting of the genes encoding acid invertase, UDP-glucose pyrophosphorylase, polyphenol oxidase, O-methyl transferase, cinnamyl alcohol dehydrogenase, *etr-1* or a homolog thereof, ACC synthase and ACC oxidase.
37. The method of claim 36, wherein the plant cell is a maize, wheat, rice or lettuce cell.
38. The method of claim 36, wherein the plant cell is a potato, tomato, canola, soybean or cotton cell.
39. The method of claim 32, which further comprises culturing the identified cell such that a plant is generated.
40. The method of claim 39, which further comprises making seeds from the plant or from progeny of the plant.
41. The method of claim 31, wherein the second recombinagenic oligonucleobase is a heteroduplex recombinagenic oligonucleobase and each of the homologous regions of the second recombinagenic oligonucleobase contains an RNA segment of at least 6 RNA-Type nucleotides.
42. The method of claim 41, wherein the first target gene is a first ALS gene, a second ALS gene, a *psbA* gene, a threonine dehydratase gene, a dihydrodipicolinate synthase gene, or an *S14/rp59* gene.
43. The method of claim 42, wherein the plant cell is a maize, wheat, rice or lettuce cell.
44. The method of claim 42, wherein the plant cell is a potato, tomato, canola, soybean or cotton cell.
45. The method of claim 41, wherein the second target gene is selected from the group consisting of the genes encoding acid invertase, UDP-glucose pyrophosphorylase, polyphenol oxidase, O-methyl transferase, cinnamyl alcohol dehydrogenase, *etr-1* or a homolog thereof, ACC synthase and ACC oxidase..
46. The method of claim 36, 45, wherein the second target gene is from a maize, wheat, rice or lettuce plant.

47. The method of claim 36, 45, wherein the second target gene is from a potato, tomato, canola, soybean or cotton plant.
48. The method of claim 41, which further comprises culturing the identified cell such that a plant is generated.
49. The method of claim 48, which further comprises making seeds from the plant or from progeny of the plant.
50. A method of making a localized mutation in a target gene in a plant cell comprising the steps of:
 - a. digesting a plant part with cellulase such that plant cell protoplasts are formed;
 - b. suspending the protoplasts in a solution comprising a recombinagenic oligonucleobase which contains a first homologous region which has a sequence identical to the sequence of at least 6 base pairs of a first fragment of the target gene and a second homologous region which has a sequence identical to the sequence of at least 6 base pairs of a second fragment of the target gene, and an intervening region which contains at least 1 nucleobase heterologous to the target gene, which intervening region connects the first homologous region and the second homologous region;
 - c. electroporating the suspension such that the recombinagenic oligonucleobase enters a protoplast of the suspension;
 - d. culturing the protoplast; and
 - e. identifying a progeny of the protoplast having a mutation located between the first and second fragments of the target gene.
51. The method of claim 50, which further comprises the step of culturing the identified progeny such that a plant is generated.
52. The method of claim 50, wherein the recombinagenic oligonucleobase is a MDON and each of the homologous regions contains an RNA segment of at least 6 RNA-Type nucleotides.

53. The method of claim 50, wherein the recombinagenic oligonucleobase is an heteroduplex recombinagenic oligonucleobase.
54. A plant or seed having a point mutation in a gene is in its wild type genetic position, which gene is selected from the group consisting of the genes encoding acid invertase, UDP-glucose pyrophosphorylase, polyphenol oxidase, O-methyl transferase, cinnamyl alcohol dehydrogenase, ACC synthase and ACC oxidase or *etr-1* or a homolog of *etr-1*, and the sequence of the genomic DNA within 23 KB of the mutation is the sequence of the wild type DNA, and the point mutation forms a stop codon or is a frameshift mutation.
55. The plant or seed of claim 54, in which the point mutation forms a stop codon.
56. The plant or seed of claim 55, in which the sequence of the genomic DNA within 40 KB of the selectable mutation is the sequence of the wild type DNA.
57. The plant or seed of claim 55, in which the sequence of the genomic DNA within 100 KB of the selectable mutation is the sequence of the wild type DNA.
58. The plant or seed of claim 55, in which the point mutation is a single base pair mutation.
59. The plant or seed of claim 55, which is a maize, wheat, rice or lettuce plant or seed.
60. The plant or seed of claim 55, which is a potato, tomato, canola, soybean or cotton plant or seed.
61. The plant or seed of claim 55, further having a selectable point mutation in a second gene and the sequence of the genomic DNA within 23 KB of the selectable point mutation is the sequence of the wild type DNA.
62. The plant or seed of claim 61, in which the sequence of the genomic DNA within 40 KB of the selectable mutation is the sequence of the wild type DNA.
63. The plant or seed of claim 61, in which the sequence of the genomic DNA within 100 KB of the selectable mutation is the sequence of the wild type DNA.
64. The plant or seed of claim 54, in which the point mutation is a frameshift

mutation.

65. The plant or seed of claim 64, in which the sequence of the genomic DNA within 40 KB of the selectable mutation is the sequence of the wild type DNA.
66. The plant or seed of claim 64, in which the sequence of the genomic DNA within 100 KB of the selectable mutation is the sequence of the wild type DNA.
67. The plant or seed of claim 64, in which the point mutation is a single base pair mutation.
68. The plant or seed of claim 64, which is a maize, wheat, rice or lettuce plant or seed.
69. The plant or seed of claim 64, which is a potato, tomato, canola, soybean or cotton plant or seed.
70. The plant or seed of claim 64, further having a selectable point mutation in a second gene and the sequence of the genomic DNA within 23 KB of the selectable point mutation is the sequence of the wild type DNA.
71. The plant or seed of claim 70, in which the sequence of the genomic DNA within 40 KB of the selectable mutation is the sequence of the wild type DNA.
72. The plant or seed of claim 70, in which the sequence of the genomic DNA within 100 KB of the selectable mutation is the sequence of the wild type DNA.

ABSTRACT

The invention concerns the use of duplex oligonucleotides about 25 to 30 base pairs to introduce site specific genetic alterations in plant cells. The oligonucleotides can be delivered by mechanical (biolistic) systems or by electroporation of plant protoplasts.

Thereafter plants having the genetic alteration can be generated from the altered cells. In specific embodiments the invention concerns alteration in the gene that encode acid invertase, UDP-glucose pyrophosphorylase, polyphenol oxidase, O-methyl transferase, cinnamyl alcohol dehydrogenase, ACC synthase and ACC oxidase or *etr-1* or a homolog of *etr-1*, and plants having isolated point mutations in such genes.

SEQUENCE LISTING

<110> 1. Arntzen, Charles
2. Kipp, Peter B.
3. Kumar, Ramesh
4. May, Gregory D.

<120> The Use of Mixed Duplex Oligonucleotides
to Effect Localized Genetic Changes in Plants

<130> 7991-023-999

<150> 60/054,386

<151> 1997-08-05

<160> 19

<170> FastSEQ for Windows Version 3.0

<210> 1

<211> 2063

<212> DNA

<213> Solanum tuberosum

<220>

<221> CDS

<222> (3) ... (1907)

<400> 1

agtagcattc	cagttatgac	ccggaaaact	ccgcctccca	ttacacattc	ctcccggatc	60
aaccggatc	cggccaccg	aagtcctta	aatcatctc	cggcattttc	ctctcctctt	120
tccttttgct	ttctgtagc	ttctttccga	tcctcaacaa	ccagtcaccg	gacttgcaga	180
gtaactcccg	ttcgcccgcc	ccgtcaagag	gtgtttctca	gggagtctcc	gataagactt	240
ttcgagatgt	cgtcaatgct	agtcacattt	cttatgcgtg	gtccaatgct	atgcttagct	300
ggcaaaagac	tgcttaccat	tttcaacctc	aaaaaaattg	gatgaacgat	cctaattggtc	360
cattgtacca	caagggatgg	tatcatcttt	tttatcaata	caatccagat	tcagctatctt	420
ggggaaatat	cacatggggc	catgccgtat	ccaaggactt	gatccactgg	ctctacttgc	480
cttttgccat	ggttcctgat	caatgggtacg	atattaacgg	tgtctggact	gggtccgcct	540
ccatcctacc	cgatggtcag	atcatgatgc	tttataccgg	tgtctctgat	gattatgtac	600
aagtgcacaa	tcttgcgtag	cccaccaact	tatctgatcc	tctccttcta	gactgggtca	660
agtacaaaagg	caaccgggtt	ctggttcctc	caccggcat	tggtatcaag	gacttttagag	720
acccgaccac	tgcttggacc	ggaccccaaa	atgggcaatg	gcttttaaca	atcgggtcta	780
agattggtaa	aacgggtatt	gcacttggtt	atgaaacttc	caacttcaca	agctttaagc	840
tattggatga	agtgtctgat	gcggttcocg	gtacgggtat	gtgggagtgt	gtggactttt	900
acccggtatc	gactgaaaaa	acaaacgggt	tggacacatc	atataacggc	ccgggtgtaa	960
agcatgtgtt	aaaagcaagt	ttagatgaca	ataagcaaga	tcactatgct	attgggacgt	1020
atgacttgac	aaagaacaaa	tggacacccg	ataacccgga	attggattgt	ggaattgggt	1080
tgaagctgga	ttatgggaaa	tattatgcat	caaagacatt	ttatgaccgg	aagaaacaac	1140
gaagagtact	gtggggatgg	attggggaaa	ctgatagtga	atctgctgac	ctgcagaagg	1200
gatgggcatc	tgtacagagt	attccaagga	cagtgcctta	cgacaagaag	acagggacac	1260
atctacttca	gtggccagtt	gaagaaattg	aaagcttaag	agtgggtgat	cctattgtta	1320
agcaagtcaa	tcttcaacca	ggttcaattg	agctactcca	tgttgactca	gctgcagagt	1380
tggaatata	agcctcattt	gaagtggaca	aagtcgcgct	ccagggaata	attgaagcag	1440

atcatgtagg	tttcagctgc	tctactagt	gaggtgctgc	tagcagaggg	atcttgggac	1500
catttgggtg	cgttgtaatt	gctgatcaaa	agctatctga	gctaacgcca	gtttacttct	1560
acatttctaa	aggagctgat	ggtcgagctg	agactcactt	ctgtgctgat	caaactagat	1620
cctcagaggg	tccgggagtt	gctaaacaag	tttatggtag	ttcagtagcc	gtgttagacg	1680
gtgaaaaaca	ttcgatgaga	ttattggagg	accactcaat	tgtggagagc	tttgcccaag	1740
gaggaagaac	agtcataaca	tcgcgaattt	acccaacaaa	ggcagtgaat	ggagcagcac	1800
gactcttcgt	tttcaacaat	gccacagggg	ctagcgtgac	tgcttccgtc	aagatttggt	1860
cacttgagtc	ggctaataat	cgatccttcc	ccttgcaaga	cttgtaattc	atcaagccat	1920
atcttcttca	ttcttttttt	catttgaagg	ttatttcacc	gatgtcccat	caagaaaggg	1980
aagagagggg	gaatatgtag	tggtatactc	tacttattcg	ccatttttagt	gatttttcta	2040
ctggactttt	gctattcgca	aaa				2063

<210> 2
 <211> 1958
 <212> DNA
 <213> Solanum tuberosum

<220>
 <221> CDS
 <222> (22)...(1815)

<400> 2						
tcttttgcgt	tttgagcaat	aatggcaagc	ttgtgcaata	gtagtagtac	atctctcaaa	60
actcctttta	cttcttcctc	cacttcttta	tcttccactc	ctaagccctc	tcaacttttc	120
atccatggaa	aacgtaacca	aatgttcaaa	gtttcatgca	aggttaccaa	taataacggt	180
gacccaaaac	aaaacgttga	aacaaattct	gttgatcgaa	gaaatgttct	tcttggctta	240
gggtggtcttt	atggtgttgc	taatgctata	ccattagctg	catccgctgc	tccagctcca	300
cctcctgata	tctcgtcttg	tagtatagcc	aggattaacg	aaaatcaggt	ggtgccgtac	360
agttgttgcg	cgcctaagcc	tgatgatatg	gagaaagttc	cgtattacaa	gttcccttct	420
atgactaagc	tccgtgttcg	tcagcctgct	catgaagcta	atgaggagta	tattgccaaag	480
tacaatctgg	cgattagtcg	aatgaaagat	cttgataaga	cacaaccttt	aaaccttatt	540
ggttttaagc	aacaagctaa	tatacattgt	gcttattgta	acggtgctta	tagaattgggt	600
ggcaaagagt	tacaagttca	taattcttgg	cttttcttcc	cgttccatag	atggtagcttg	660
tacttccacg	agagaatcgt	gggaaaattc	attgatgata	caactttcgc	tttaccatat	720
tggaattggg	accatccaaa	aggtatgcgt	tttccctgca	tgtagatcg	tgaagggact	780
tcccttttctg	atgtaacacg	tgaccaaagt	caccgaaatg	gagcagtaat	cgatcttgggt	840
tttttcggca	atgaagttga	aacaactcaa	ctccagttga	tgagcaataa	tttaacacta	900
atgtaccgtc	aatggtaaac	taatgctcca	tgctctcgga	tggtcttttg	cgggccttat	960
gatctcgggg	ttaacactga	actcccggga	actatagaaa	acatccctca	cggtcctgtc	1020
cacatctggg	ctggtacagt	gagaggttca	actttgcccc	atgggtgcaat	atcaaacggt	1080
gagaatatgg	gtcattttta	ctcagctggg	ttggaccggg	ttttcttttg	ccatcacagc	1140
aatgtggatc	ggatgtggag	cgaatggaaa	gcgacaggag	ggaaaagaac	ggatatcaca	1200
cataaagatt	ggttgaactc	cgagttcttt	ttctatgatg	aaaatgaaaa	cccttaccgt	1260
gtgaaagtca	gagactgttt	ggacacgaag	aagatgggat	acgattacaa	accaattgcc	1320
acaccatggc	gtaacttcaa	gcccttaaca	aaggcttcag	ctggaaaagt	gaatacagct	1380
tcacttccgc	cagctagcaa	tgtattccca	ttggctaaac	tcgacaaagc	aatttcgttt	1440
tccatcaata	ggccgacttc	gtcaaggact	caacaagaga	aaaatgcaca	agaggagatg	1500
ttgacattca	gtagcataag	atatgataac	agagggtaca	taagggttcga	tgtgttttcg	1560
aacgtgggac	ataatgtgaa	tgcgaaatgag	cttgacaagg	cggagtttgc	ggggagttat	1620
acaagtttgc	cacatgttca	tagagctggg	gagactaatc	atatcgcgac	tggttgatttc	1680
cagctggcga	taacgggaact	gttggaggat	attgggtttg	aagatgaaga	tactattgcg	1740
gtgactctgg	tgccaaagag	aggtgggtgaa	ggtatctcca	ttgaagggtgc	gacgatcagt	1800
cttgacagatt	gttaattagt	ctctattgaa	tctgctgaga	ttacactttg	atggatgatg	1860
ctctgttttt	gttttcttgt	tctgtttttt	cctctgttga	aatcagcttt	gttgcttgat	1920
ttcattgaag	ttgttattca	agaataaata	agttacaa			1958

<210> 3
 <211> 1460
 <212> DNA
 <213> Nicotiana tabacum

<220>
 <221> CDS
 <222> (84)...(1178)

<400> 3

tctgtttctt	caactcacct	taatttgccc	aattgagtc	ttgtaaaatc	tgaacacagaa	60
ccaagagaga	agagaaaaaa	aatatgggtt	caacaagcca	gagccagagt	aagagtctaa	120
ctcacacaga	agacgaagcg	ttcttatttg	ccatgcaatt	ggctagtgtt	tctgtacttc	180
ctatggctct	aaaatcagcg	ttagaacttg	accttcttga	actcatggct	aaagctgggc	240
caggtgcagc	catttctcct	tctgaattag	ctgctcagct	ctcaaccag	aaccagaaag	300
caccggttat	tcttgatcgg	atgcttaggc	tacttgctac	ttactctgtt	ctcaattgta	360
ctcttagaac	actgtctgat	ggcagtgttg	agaggcttta	tagtctggct	ccggtttgta	420
agttcttgac	taagaatgct	gatgggtgtt	ctgttgcccc	acttttgctt	atgaatcaag	480
ataaagtctt	tatggagagc	tggtaccact	taaaagatgc	agtactagat	ggtggaatcc	540
cattcaacaa	ggcctatgga	atgacagcat	ttgagtacca	tggcacagat	ccaagattca	600
acaaagtttt	caaccgtgga	atgtctgac	actccactat	gtcaatgaaa	aagattcttg	660
aggactacaa	aggatttgaa	ggcctaaatt	ccattgttga	tgttggtggt	ggaactggcg	720
ctactgttaa	catgattgtc	tccaaacatc	cctctattaa	gggtattaac	tttgattttac	780
cacatgttat	tggagatgct	ccagcttacc	ctgggtgtcga	gcacgttggt	ggcgacatgt	840
ttgccagtg	gccccaaagca	gatgccattt	tcatgaagt	gatttgcac	gattggagcg	900
acgagcattg	cctaaaattc	ttgaagaatt	gctatgaagc	actacctgca	aatgggaagg	960
tgataatagc	ggagtgcata	cttccagagg	ccccagatac	atcacttgca	actaagaata	1020
cagtacatgt	tgatattgtg	atgttagcac	ataaccagag	aggcaaagaa	aggactgaga	1080
aggaatttga	ggcctttggc	aaggcgctg	gttttactgg	attcgcaagg	cttgttgccg	1140
ttacaacact	tgggtcatgg	aattcaacaa	ataattaatc	gattcctttg	gaggattaag	1200
caatatactg	ttcattttgc	attttgaaat	tctacttttc	acagagtggc	tttactgcga	1260
aataaaagaa	atatatagct	tttaccttga	aaagatcaat	gttcaaagg	aaaaaaaaaa	1320
aggaagatga	aataattgct	ctcagaaaag	cagtgtgtta	ggaaaaagct	ttttagctgg	1380
attttgaatt	ttattgtatg	tatttctgta	atacacatgt	attgaaggaa	tactagtttt	1440
cgaccaatca	tatttctttg					1460

<210> 4
 <211> 1418
 <212> DNA
 <213> Nicotiana tabacum

<220>
 <221> CDS
 <222> (59)...(1153)

<400> 4

attccttcaa	cttaccat	taagtcacg	aaaaatctga	aacagaacta	aaagtaaaat	60
gggttcaaca	agcgagagcc	agagtaacag	tctactcac	acagaagacg	aagctttctt	120
atttgccatg	caattgtgta	gtgcttctgt	acttccatg	gtcctaaaat	cagccgtaga	180
acttgacctt	cttgagctaa	tggttaaggc	tggtccaggt	gcagctat	ctccttctga	240
attagctgct	cagctctcaa	ctcagaacct	agaagcacct	gttatgctt	atcggtatgct	300
taggtactt	gcttcttact	ctgttctcaa	ttgtactctt	agaacactgc	ctgatagcag	360
tggtgagagg	ctttatagtc	tggtcccggt	ctgtaagtac	ttgactaaga	atgctgatgg	420
tggttctgtt	gccccacttt	tgcttatgaa	tcaagataaa	gttcttatgg	agagctggta	480
ccacttaaaa	gatgcagtac	tagatggcgg	aatcccatc	aacaaagcct	atggaatgac	540

agcatttgag	taccatggca	cagatccaag	attcaacaaa	gtgttcaacc	gtggaatgtc	600
tgatcactcc	actatgtcaa	tgaagaagat	tcttgaggac	tacaaaggat	ttgaaggcct	660
aaattccatt	gttgatgttg	gtggtggaac	gggtgctact	gttaacatga	ttgtctctaa	720
atatccctct	attaagggca	ttactttga	tttgccacat	gtaattggag	atgctccaac	780
ttaccccggt	gtcgagcacg	ttggtggcga	catgtttgct	agtgtgcaa	aagcagatgc	840
cattttcatg	aagtggattt	gtcatgattg	gagcgatgag	cattgcctaa	aattcttgaa	900
gaattgctat	gaagcactac	ctgcaaatgg	gaagggtgata	attgcagagt	gcatacttcc	960
agaggcccca	gatacatcac	ttgcaactaa	gaatacagta	catgttgata	ttgttatgtt	1020
agcacataac	ccaggaggca	aagaaaggac	tgagaaggaa	tttgaggctt	tggctaaggg	1080
cgctggtttt	actggattcg	caaggcttgt	tgcgcttaca	acacttgggt	catggaattc	1140
aacaagtaat	taatcgattc	cttaatttga	aggattaagc	aatatactgt	tcgttttgca	1200
tttgaaaatt	ctacttttct	cagagtggct	tgactgtgaa	ataaaagaaa	tatagctttt	1260
aacttgaaaa	gattgatgtt	caaaagaaaa	aaaggaagat	gaaataattg	ctctcagaaa	1320
agcaatgtgt	taggaaaaag	cttttttagc	tggattttga	attttactgt	atgtatttct	1380
gttatacaca	tgtattgaag	gaatactagt	tttcgacc			1418

<210> 5
 <211> 1419
 <212> DNA
 <213> Nicotiana tabacum

<220>
 <221> CDS
 <222> (92)...(1165)

<400> 5						
atttctttct	ctttcccttg	aactgtgttt	tcattttttc	tgctctgaaa	caatagtgtt	60
ttcctttag	attttaagtt	aaaagaaaac	catgggtagc	ttggatgttg	aaaaatcagc	120
tattggttgg	gctgctagag	acccttctgg	tctactttca	ccttatacct	atactctcag	180
aaacacagga	cctgaagatg	tgcaagtcaa	agttttgtat	tgtggacttt	gccacagtga	240
tcttcaccaa	gttaaaaatg	atcttggcat	gtccaactac	cctctggttc	ctggacatga	300
agtgggtgga	aaagtagtgg	aggtaggagc	agatgtgtca	aaattcaaag	tgggggacac	360
agttggagtt	ggattactcg	ttggaagtgg	taggaactgt	ggcccttgca	agagagaaat	420
agagcaatat	tgcaacaaga	agatttggaa	ttgcaatgat	gtctacactg	atggcaaacc	480
caccaaggt	ggttttgcta	attctatggt	tgttgatcaa	aactttgtgg	tgaaaattcc	540
agagggtatg	gcaccagaac	aagcagcacc	tctattatgt	gctggcataa	cagtatacag	600
tccattcaac	cattttgggt	ttaatcagag	tggatttaga	ggaggaattt	tgggattagg	660
aggagttgga	catatgggag	tgaaaatagc	aaaggcaatg	ggacatcatg	ttactgtcat	720
tagttcttca	aataagaaga	gacaagaggc	attggaacat	cttgggtgcag	atgattatct	780
tgtagtttca	gacactgata	aaatgcaaga	agctgctgat	tcacttgact	atattattga	840
tactgtccct	gttggccatc	ctcttgaact	ttatctttct	ttgcttaaaa	ttgatggcaa	900
acttatcttg	atcggagttt	tcaacacccc	cttgcaattt	atctctccca	tggttatgct	960
cgggagaaaag	agcatcactg	gaagctttat	tggttagcatg	aaggaaacag	aggaaatgct	1020
agacttctgc	aaagagaaaag	gtgtgacttc	acagatttag	atagtgaaaa	tggattatat	1080
caacactgca	atggagaggt	tggagaaaaa	tgatgtgagc	tacagatttg	ttgttgatgt	1140
tgctggaagc	aagcttgacc	agtaattgca	caagaaaaac	aacatggaat	ggttcaactat	1200
tatacaacaa	ggctatgaga	aaaatagtac	tcctcaactt	tgatgtcatc	tttgttacct	1260
ttgttttatt	ttccacctgt	attatcatat	ttgggtggctg	agagtgcagt	ttatgtatat	1320
tttctttctt	caaaacaatc	ttaaatgaat	ttggatgttg	gtgacgattt	tgaaatatac	1380
caaccatgca	aacttacttt	ggtagaaaaa	aaaaaaaaa			1419

<210> 6
 <211> 1398
 <212> DNA
 <213> Nicotiana tabacum

<220>
 <221> CDS
 <222> (88)...(1161)

<400> 6

attcctcttt	cccttgaact	gtgttttcgt	tttttctgct	ctaaaacaat	cgtgtgttcc	60
ttctagattt	taagttttaa	gaacatcatg	ggtggcttgg	aagttgagaa	aacaactatt	120
ggttgggctg	ctagagaccc	ttctggtgta	ctttcacctt	atacctatac	tctcagaaac	180
acaggacctg	aagatgtgga	agtcaaagtt	ttgtattgtg	ggctctgtca	cactgatctt	240
caccaagtta	aaaatgatct	tggcatgtcc	aactaccctc	tggttcctgg	acatgaagtg	300
gtgggagaa	tggtggaggt	aggaccagat	gtgtcaaaat	tcaaagttgg	ggacacagtt	360
ggagttggat	tactcgttgg	aagttgcagg	aactgtggcc	cttgcaagag	agatatagag	420
caatattgca	acaagaagat	ttggaactgc	aatgatgtct	acactgatgg	caaaccacc	480
caaggtggtt	ttgctaaatc	catggttgtt	gatcaaaagt	ttgtggtgaa	aattccagag	540
ggtatggcac	cagaacaagc	agcacctcta	ttatgtgctg	gtataacagt	atacagtcca	600
ttgaaccatt	ttggtttcaa	acagagtgga	ttaagaggag	gaattttggg	attaggagga	660
gtgggacaca	tggtgagtaa	aatagcaaag	gcaatgggac	atcatgttac	tgctattagt	720
tcttcaaata	agaagagaca	agaggcattg	gaacatcttg	gtgcagatga	ttatcttgtc	780
agttcagaca	ctgataaaat	gcaagaggct	tctgattcac	ttgactatat	tattgatact	840
gtccctgttg	gccatcctct	tgaaccttat	ctttctttgc	ttaaaattga	tggcaaactt	900
atcttgatgg	gagttatcaa	caccccttg	caatttatct	cccccatggt	tatgctcggg	960
agaaagagca	tcacaggaag	ctttatttgg	agcatgaagg	aaacagagga	aatgctagat	1020
ttctgcaaag	agaaaggtgt	gacttcacag	attgagatag	tgaaaatgga	ttatatcaac	1080
actgcaatgg	agaggttggg	gaaaaatgat	gtgaggtaca	gattttgtgt	tgatgttatt	1140
ggaagcaagc	ttgaccagta	attatattac	acaagaaaaa	caacatggaa	tggttcacta	1200
ttatacaagg	ctgtgagaat	actaaacttt	gatgtcgtct	tttgtatcct	tttgttttat	1260
ttgccacctg	tattttctta	tttggtgatc	gagagtgcgc	tttatgtatt	attttctttc	1320
ttcaaaacaa	tttaatgtat	gaatttggat	gttgggtgaaa	aaaaaaaaaa	aaaaaaaaaa	1380
aaaaaaaaaa	aaaaaaaaaa					1398

<210> 7
 <211> 1533
 <212> DNA
 <213> *Carthamus tinctorius*

<220>
 <221> CDS
 <222> (106)...(1296)

<400> 7

gctcacttgt	gtggtggagg	agaaaaacag	aactcacaaa	aagcttttgcg	actgccaaga	60
acaacaacaa	caacaagatc	aagaagaaga	agaagaagat	caaaaatggc	tcttcgaatc	120
actccagtga	ccttgcaatc	ggagagatat	cgttcgtttt	cgtttcctaa	gaaggctaata	180
ctcagatctc	ccaaatfcgc	catggcctcc	accctcggat	catccacacc	gaaggttgac	240
aatgccaaga	agccttttca	acctccacga	gaggttcatg	ttcagggtgac	gcaactccatg	300
ccaccacaga	agatagagat	tttcaaattc	atcgagggtt	gggctgagca	gaacatattg	360
gttcacctaa	agccagtggg	gaaatgttgg	caagcacagg	atttcttgcc	ggaccctgca	420
tctgaaggat	ttgatgaaca	agtcaaggaa	gagcaaagga	gattcctgat		480
gattactttg	ttgttttggg	tggagatatg	attacagagg	aagccctacc	tacttaccaa	540
acaatgctta	ataccctaga	tggtgtacgt	gatgagactg	gggctagcct	tacgccttgg	600
gctgtctgga	ctagggcttg	gacagctgaa	gagaacaggc	atggcgatct	tctccacacc	660
tatctctacc	tttctgggcg	ggtagacatg	aggcagatac	agaagacaat	tcagtatctc	720
attgggtcag	gaatggatcc	tcgtaccgaa	aacagccctc	accttgggtt	catctacaca	780
tcgtttcaag	agcgtgccac	atttgtttct	cacggaaaca	ccgccaggca	tgcaaaggat	840
catggggacg	tgaaactggc	gcaaatttgt	ggtacaatcg	cgtctgacga	aaagcgtcac	900

gagaccgctt	atacaaaagat	agtcgaaaaag	ctattcgaga	tcgatccctga	tggcaccggt	960
cttgctttttg	ccgacatgat	gaggaaaaaag	atctcgatgc	ccgcacactt	gatgtacgat	1020
gggcgtgatg	acaacctctt	cgaacatttc	tcggcggttg	cccaaagact	cggcgtctac	1080
accgccaaaag	actacgccga	catactggaa	tttctggctg	ggcggtggaa	agtggcggat	1140
ttgaccggcc	tatctggtga	agggcgtaaa	gcgcaagatt	atgtttgctg	gttgccacca	1200
agaatcagaa	ggctggagga	gagagctcaa	ggcgagcaa	aggaaggacc	tgttgttcca	1260
ttcagctgga	ttttcgatag	acaggtgaag	ctgtgaagaa	aaaaaaaacg	agcagtgaag	1320
tcggtttctg	ttggcttatt	gggtagaggt	taaaacctat	tttagatgtc	tgtttcgtgt	1380
aatgtggttt	tttttcttct	aatcttgaat	ctggtattgt	gtcgttgagt	tcgcgtgtgt	1440
gtaaacttgt	gtggctgtgg	acatattata	gaactcgtaa	tgccaatttt	gatgacggtg	1500
gttatcgtct	ccctcggtgt	ttttttattg	ttt			1533

<210> 8
 <211> 1643
 <212> DNA
 <213> Ricinus communis

<220>
 <221> CDS
 <222> (1)... (1239)

ttccggcaaa	taacaaaaaa	ccaaaagaaa	aaggtaagaa	aaaaaacaat	ggctctcaag	60
ctcaatcctt	tccttttctca	aacccaaaaag	ttaccttctt	tcgctcttcc	accaatggcc	120
agtaccagat	ctoctaagtt	ctacatggcc	tctacctca	agtctgggtc	taaggaagtt	180
gagaatctca	agaagccttt	catgcctcct	cgggaggtac	atgttcaggt	taccattct	240
atgccacccc	aaaagattga	gatctttaaa	tccctagaca	attgggctga	ggagaacatt	300
ctggttcac	tgaagccagt	tgagaaatgt	tggcaaccgc	aggatTTTTT	gccagatccc	360
gcctctgatg	gatttgatga	gcaagtcagg	gaactcagg	agagagcaaa	ggagattcct	420
gatgattatt	ttgttggttt	ggttggagac	atgataacgg	aagaagccct	tcccacttat	480
caaacaatgc	tgaatacctt	ggatggagtt	cgggatgaaa	caggtgcaag	tcctacttct	540
tgggcaattt	ggacaagggc	atggactcgc	gaagagaata	gacatggtga	cctcctcaat	600
aagtatctct	acctatctgg	acgagtggac	atgaggcaaa	ttgagaagac	aattcaatat	660
ttgattgggt	caggaatgga	tccacggaca	gaaaacagtc	cataccttgg	gttcatctat	720
acatcattcc	aggaaagggc	aaccttcatt	tctcatggga	acactgccc	acaagccaaa	780
gagcatggag	acataaagtt	ggctcaaata	tgtggtacaa	ttgctgcaga	tgagaagcgc	840
catgagacag	cctacacaaa	gatagtggaa	aaactctttg	agattgatcc	tgatggaact	900
gttttggtct	ttgctgatat	gatgagaaa	aaaatttcta	tgctgcaca	cttgatgtat	960
gatggccgag	atgataatct	ttttgaccac	ttttcagctg	ttgctgcagc	tcttggagtc	1020
tacacagcaa	aggattatgc	agatatattg	gagttcttgg	tgggcagatg	gaagggtggat	1080
aaactaacgg	gcctttcagc	tgagggacaa	aaggctcagg	actatgtttg	tcggttacct	1140
ccaagaatta	gaaggctgga	agagagagct	caaggaagg	caaaggaagc	acccaccatg	1200
cctttcagct	ggattttcga	taggcaagtg	aagctgtagg	tggctaaaag	gcaggacgaa	1260
accgaaatgg	ttagtttcac	tctttttcat	gccatccct	gcagaatcag	aagtagaggt	1320
agaattttgt	agttgctttt	ttattacaag	tccagtttag	tttaaggtct	gtggaaggga	1380
gttagttgag	gagtgaatgt	agtaagttgt	tgatactgtt	gtgttcttgt	gttgatcatga	1440
gtctgcttga	tagtgagttt	cttttggttc	cttttggtgt	gttcttttat	ctggtctctc	1500
tctctctctc	tctctctttt	tctcttatcc	caagtgtctc	aagtataata	agcaaacgat	1560
ccatgtggca	atthttgatga	tggtgatcag	tctcacaact	tgatcttttg	tcttctattg	1620
gaaacacagc	ctgcttggtt	gaa				1643

<210> 9
 <211> 2569
 <212> DNA
 <213> Arabidopsis thaliana

<220>
 <221> exon
 <222> (236) ... (729)
 <223> Exon 1

 <221> exon
 <222> (1030) ... (1119)
 <223> Exon 2

 <221> exon
 <222> (1201) ... (1267)
 <223> Exon 3

 <221> exon
 <222> (1358) ... (1450)
 <223> Exon 4

 <221> exon
 <222> (1530) ... (1715)
 <223> Exon 5

 <221> exon
 <222> (1809) ... (1889)
 <223> Exon 6

 <221> exon
 <222> (1993) ... (2130)
 <223> Exon 7

 <221> exon
 <222> (2212) ... (2403)
 <223> Exon 8

<400> 9

cacaccatca	ctaataaatt	tccttctcct	ttcaagttgt	agctaactta	tataagacat	60
aagcgtgcga	accagagaca	gagatagaaa	ttgagagacg	ataagcaaag	tagaaaacac	120
aagtttctct	cacacacatt	atctctttct	ctattaccac	cactcattca	taacagaaac	180
ccacaaaaaa	ataaaaagag	agacttttca	ctctggggag	agagctcaag	ttctaattggc	240
gaacttggtc	ttatcagaat	gtggtatacg	acctctcccc	agaatctaca	caacacccag	300
atccaatttc	ctctccaaca	acaacaaatt	cagaccatca	ctttcttctt	cttcttacia	360
aacatcatca	tctcctctgt	cttttggtct	gaattcacga	gatgggttca	cgaggaattg	420
ggcgttgaat	gtgagcacac	cattaacgac	accaatattt	gaggagtctc	cattggagga	480
agataataaa	cagagattcg	atccaggtgc	gcctcctccg	ttcaatttag	ctgatattag	540
agcagctata	cctaagcatt	gttgggttaa	gaatccatgg	aagtctttga	gttatgtcgt	600
cagagacgtc	gctatcgtct	ttgcattggc	tgctggagct	gcttacctca	acaattggat	660
tgtttggect	ctctattggc	tcgctcaagg	aaccatgttt	tgggctctct	ttgttcttgg	720
tcattgactgg	taaacttaaa	aaccctaact	tttttcttgt	tttctcctct	gctttagtct	780
cctttagcct	ttgatttggg	caactttgga	tgattccaaa	gaaccaatcg	aacaaattgg	840
tctttatcca	tatctcttca	aatagcttta	ggacataaatt	ggtctctcag	gtaacaagct	900
gtcattatca	tcatactcat	catgttgcta	gtagaccaac	ccaattggca	actggttgtt	960
ggttttgcaa	ctgtgtaatc	tgctttgaat	tgtgaacaaa	attattgatt	tatgttgatt	1020
acattgcagt	ggacatggta	gtttctcaaa	tgatccgaag	ttgaacagtg	tggtcgggtca	1080
tcttcttcat	tcctcaattc	tggtcccata	ccatggctgg	tgagttttgc	tttcagacca	1140
ttcttctcta	aaaccacttg	cagaatctca	tcttcttcat	gtaaaaatat	gactttgcag	1200
gagaattagt	cacagaactc	accaccagaa	ccatggacat	gttgagaatg	acgaatcttg	1260

gcacccctgta	agtcaaaaaac	gtattttttt	ggttatcttg	ttttagtcct	gtgggtgttc	1320
ttagatgcag	ttttattaac	tgtttctgta	actgcagatg	tctgagaaaa	tctacaatac	1380
tttgacaag	ccgactagat	tcttttagatt	tacactgcct	ctcgtgatgc	ttgcataccc	1440
tttctacttg	gtaagaactc	ctctatttgt	tatggtaact	taagctgcca	caccaagtaa	1500
aaaagctcat	gtctattctt	ctgtttcagt	gggctcgaag	tccggggaaa	aagggttctc	1560
attaccatcc	agacagtgac	ttgttctctc	ctaaagagag	aaaggatgtc	ctcacttcta	1620
ctgcttggtg	gactgcaatg	gctgctctgc	ttgtttgtct	caacttcaca	atcgggtccaa	1680
ttcaaatgct	caaactttat	ggaattcctt	actgggtaat	gcgccgctgt	tactcccttg	1740
tttcagcctg	agcaatttgt	gtattatttc	ctctgcctta	ctcaaaaagg	tttttatgtc	1800
aaatacagat	aatgtaatg	tggttggact	ttgtgactta	cctgcatcac	catgggtcatg	1860
aagataagct	tccttggtac	cgtggcaagg	taaaatacat	attctctgct	tccactgttc	1920
tttgactaca	tcgtcttttc	ttttaagggt	aaagccaact	ggtgtgtaaa	tctcatgatt	1980
ctcccaaaac	aggagtggag	ttacctgaga	ggaggactta	caacattgga	tcgtgactac	2040
ggattgatca	ataacatcca	tcatgatatt	ggaactcatg	tgatacatca	tcttttcccg	2100
cagatcccac	attatcatct	agtagaagca	gtaagtaaat	tgaaagtaaa	gactgtttgt	2160
gtttttggtg	ttcatgctag	tttcctgac	tcttgctcca	ctgttatgca	gacagaagca	2220
gctaaaccag	tattagggaa	gtattacagg	gagcctgata	agtctggacc	gttgccatta	2280
catttactgg	aaattctagc	gaaaagtata	aaagaagatc	attacgtgag	cgacgaagga	2340
gaagttgtat	actataaagc	agatccaaat	ctctatggag	aggtcaaagt	aagagcagat	2400
tgaaatgaag	caggcttgag	attgaagttt	tttctatttc	agaccagctg	attttttgct	2460
tactgtatca	atttattgtg	tcacccacca	gagagttagt	atctctgaat	acgatcgatc	2520
agatggaaac	aacaaatttg	tttgcgatac	tgaagctata	tataccata		2569

<210> 10
 <211> 3879
 <212> DNA
 <213> Arabidopsis thaliana

<220>
 <221> exon
 <222> (780)...(1685)
 <223> Exon 1

 <221> exon
 <222> (1761)...(2129)
 <223> Exon 2

<221> exon
 <222> (2207)...(2461)
 <223> Exon 3

<221> exon
 <222> (2544)...(2671)
 <223> Exon 4

<221> exon
 <222> (2762)...(2959)
 <223> Exon 5

<221> exon
 <222> (3088)...(3448)
 <223> Exon 6

<400> 10
 aaagatagta tttgttgata aatatgggga tatttatcct atattatctg tatttttctt

60

accattttta	ctctattcct	ttatctacat	tacgtcatta	cactatcata	agatatttga	120
atgaacaaat	tcatgcaccc	accagctata	ttaccctttt	ttattaaaaa	aaaacatctg	180
ataataataa	caaaaaaatt	agagaaatga	cgtcgaaaaa	aaaagtaaga	acgaagaaga	240
agtgttaaac	ccaaccaatt	ttgacttgaa	aaaaagcttc	aacgctcccc	ttttctcctt	300
ctccgtcgct	ctccgcgcgc	tcccaaattcc	ccaattcctc	ctcttctccg	atcaattcctt	360
cccaagtaag	cttcttcttc	ctcgattctc	tccctcagatt	gtttcgtgac	ttctttatat	420
atattcttca	cttccacagt	tttcttctgt	tgttgctgct	gatctcaaat	catagagatt	480
gattaaccta	attggtcttt	atctagtgtg	atgcacgtt	attaggaact	ttaaattaag	540
atttaatcgt	taatttcattg	attcggattc	gaattttact	gttctcgaga	ctgaaatatg	600
caacctattt	tttcgtaatc	gttgatgatc	aattcgattc	ttcagaattt	atagcaattt	660
tgatgctcat	gatctgtcta	cgctacgttc	tcgtcgtaaa	tcgaagtga	taatgctatg	720
tgtttgttac	acaggtgtgt	gtatgtgtga	gagaggaaact	atagtgtaaa	aaattcataa	780
tggaggtctg	caattgtatt	gaaccgcaat	ggccagcgga	tgaattgtta	atgaaatacc	840
aatacatctc	cgatttcttc	attgcgattg	cgtatttttc	gattcctctt	gagttgattt	900
actttgtgaa	gaaatcagcc	gtgtttccgt	atagatgggt	acttggttcag	tttgggtgctt	960
ttatcgttct	ttgtggagca	actcatctta	ttaacttatg	gactttcact	acgcattcga	1020
gaaccgtggc	gcttgtgatg	actaccgcga	aggtgttaac	cgtgtgtgct	tcgtgtgcta	1080
ctgcgttgat	gcttgttcat	attattctctg	atcttttgag	tgtaaagact	cgggagcttt	1140
tcttgaaaaa	taaagctgct	gagctcgata	gagaaatggg	attgattcga	actcaggaag	1200
aaaccggaag	gcatgtgaga	atgttgactc	atgagattag	aagcacttta	gatagacata	1260
ctattttaaa	gactacactt	gttgagcttg	gtaggacatt	agctttggag	gagtggtcat	1320
tgtggatgcc	tactagaact	gggttagagc	tacagctttc	ttatacactt	cgtcatcaac	1380
atcccgtgga	gtatacgggt	cctattcaat	taccggtgat	taaccaagtg	tttgggtacta	1440
gtagggctgt	aaaaatatct	cctaattctc	ctgtggctag	gttgagacct	gtttctggga	1500
aatatatgct	aggggaggtg	gtcgtgtgta	gggttccgct	tctccacctt	tctaattttc	1560
agattaatga	ctggcctgag	ctttcaacaa	agagatatgc	tttgatgggt	ttgatgcttc	1620
cttcagatag	tgcaaggcaa	tggcatgtcc	atgagttgga	actcgttgaa	gtcgtcgctg	1680
atcaggtttt	acattgctga	gaatttctct	tctttgctat	gttcatgac	ttgtctataa	1740
cttttcttct	cttatttatag	gtggctgtag	ctctctcaca	tgctgcgac	ctagaagagt	1800
cgatgcgagc	tagggacctt	ctcatggagc	agaattgtgc	tcttgatcta	gctagacgag	1860
aagcagaaac	agcaatccgt	gcccgcaatg	atttcttagc	ggttatgaac	catgaaatgc	1920
gaacaccgat	gcatgcgatt	attgcactct	cttccctact	ccaagaaacg	gaactaacc	1980
ctgaacaaag	actgatgggt	gaaacaatac	ttaaaagtag	taaccttttg	gcaactttga	2040
tgaatgatgt	cttagatctt	tcaaggtag	aagatggaag	tcttcaactt	gaacttggga	2100
cattcaatct	tcatacatta	tttagagagg	taacttttga	acagctctat	gtttcataag	2160
tttatactat	ttgtgtactt	gattgtcata	ttgaatcttg	ttgcagggtc	tcaatctgat	2220
aaagcctata	gcggttggtt	agaaattacc	catcacacta	aatcttgac	cagatttgcc	2280
agaatttggt	gttggggatg	agaaacggct	aatgcagata	atattaaata	tagttggtaa	2340
tgctgtgaaa	ttctccaaac	aaggtagtag	ctccgtaacc	gctcttgtea	ccaagtcaga	2400
cacacgagct	gctgactttt	ttgtcgtgcc	aaactggagt	catttctact	tgagagtga	2460
ggttattatc	ttgtatcttg	ggatcttata	ccatagctga	aagtatttct	taggtcttaa	2520
ttttgatgat	tattcaaata	taggtaaaag	actctggagc	aggaataaat	cctcaagaca	2580
ttccaaagat	tttcaactaa	tttgctcaaa	cacaatcttt	agcgacgaga	agctcgggtg	2640
gtagtgggct	tggcctcgcc	atctccaaga	ggtttgagcc	ttattaaaag	acgttttttt	2700
ccaacttttt	cttgtcttct	gtgttggtta	aagtttactc	ataagcgttt	aatatgacaa	2760
ggtttgatga	tctgatggag	ggtaacattt	ggattgagag	cgatgggtctt	ggaaaaggat	2820
gcacggctat	ctttgatgtt	aaacttggga	tctcagaacg	ttcaaacgaa	tctaaacagt	2880
cgggcatacc	gaaagtcca	gccattcccc	gacattcaaa	tttcaactgga	cttaagggttc	2940
ttgtcatgga	tgagaacggg	ttagtataag	cttctcacct	ttctctttgc	aaaatctctc	3000
gccttacttc	ttgcaaatgc	agatattggc	gtttagaaaa	aacgcaaat	taatcttatg	3060
agaaaccgat	gattattttt	gttgcagggt	aagtagaatg	gtgacgaagg	gacttcttgt	3120
acaccttggg	tgcgaagtga	ccacgggtgag	ttcaaacgag	gagtgctctc	gagttgtgtc	3180
ccatgagcac	aaagtgggtc	tcatggacgt	gtgcatgccc	ggggctgaaa	actaccaaat	3240
cgctctccgt	attcacgaga	aattcacaaa	acaacgccac	caacggccac	tacttggtggc	3300
actcagtggt	aacactgaca	aatccacaaa	agagaaatgc	atgagctttg	gtctagacgg	3360

tgtgttgctc	aaaccggtat	cactagacaa	cataagagat	gttctgtctg	atcttctcga	3420
gccccgggta	ctgtacgagg	gcatgtaaag	gcatgggatg	ccccatgcc	cagaggagta	3480
attccgctcc	cgcttcttc	tcccgtaaaa	catcggaagc	tgatgttctc	tggtttaatt	3540
gtgtacatat	cagagattgt	cggagcggtt	tggatgatat	cttaaaacag	aaagggaata	3600
acaaaataga	aactctaaac	cggtatgtgt	ccgtggcgat	ttcgggtata	gaggaacaag	3660
atggtggtgg	tataatcata	ccatttcaga	ttacatgttt	gactaatgtt	gtatccttat	3720
atatgtagtt	acattcttat	aagaatttgg	atcgagttaa	ggatgcttgt	tgcggtcatg	3780
tatgacattg	atgcagtatt	atggcgctcag	ctttgcgccg	cttagtagaa	caacaacaat	3840
ggcggtactt	agtttctcaa	tcaaccgcat	ctccaaaac			3879

<210> 11
 <211> 1200
 <212> DNA
 <213> Arabidopsis thaliana

<220>
 <221> CDS
 <222> (53)...(1024)

<400> 11						
cgttgctgtc	gaagttaggc	caagaaaccc	atttaaaaaa	aaagagagag	agatggagag	60
tttcccgatc	atcaatctcg	agaagcttaa	tggagaagag	agagcaatca	ctatggagaa	120
gatcaaagac	gcttggtgaa	actggggcct	ctttgagtgt	gtgaaccatg	ggatttcact	180
cgagcttttg	gacaaagtgg	agaagatgac	caaggaacat	tacaagaagt	gcatggaaga	240
gagattcaag	gaatcgatta	agaacagagg	tcttgactct	cttcgctctg	aagtcaacga	300
cgttgactgg	gaatccactt	tctacctcaa	gcaccttccc	gtctctaata	tctccgatgt	360
ccctgatctc	gacgacgatt	acagaacggt	aatgaaagac	ttcgccggaa	agatagagaa	420
gttgctggag	gagctactgg	atctgctgtg	cgagaatctc	ggtttagaga	agggttattt	480
aaaaaagggtg	ttttacgggt	cgaaaagacc	gacttttggg	accaaagtca	gcaattatcc	540
accttgctct	aatccggacc	tagtcaaggg	tctccgagcc	cacaccgacg	ccggcggcat	600
catcctcttc	ttccaagacg	acaaagtcag	tggacttcag	cttcttaag	acggcgagtg	660
ggtcgatggt	cctccgggta	agcattcaat	cgtcgttaat	ctcggcgatc	aacttgaggt	720
gataaccaat	gggaagtaca	agagtgtgga	acatagagtg	ctatctcaga	cagacggaga	780
aggaagaatg	tcgatcgcat	cattctataa	tccgggaagc	gactctgtta	tttttccggt	840
gccggagctg	atcggaaaaag	aagcagagaa	ggagaagaaa	gagaactatc	cgagatttgt	900
gtttgaagat	tacatgaaac	tctactctgc	tgtcaagttt	caggccaagg	aaccaagggt	960
tgaagccatg	aaagctatgg	agacaactgt	ggccaacaat	gttggaacct	tgggcactgc	1020
gtgaatgata	tgtaactggg	taataaatat	atatatatat	atatatatag	tctttatata	1080
atgtcttaga	aacttgatta	ttcactatac	gaataatttt	gttcatgttg	ttgtatgttt	1140
aagtggtgaa	tgtgttatat	atgggaatta	atgttttctg	ttcgaaaaaa	aaaaaaaaaa	1200

<210> 12
 <211> 3438
 <212> DNA
 <213> Arabidopsis thaliana

<220>
 <221> exon
 <222> (1212)...(1358)
 <223> Exon 1

<221> exon
 <222> (1461)...(1592)
 <223> Exon 2

<221> exon
 <222> (1660)... (1820)
 <223> Exon 3

<221> exon
 <222> (1909)... (2893)
 <223> Exon 4

<400> 12

gttacttttc	aaatcttccc	tcatattata	tagccattga	tatcatagag	gatgtgagtt	60
ttaacttaat	atthaccogt	ttgaaactag	ctattttactt	aaatatgaat	tataatctag	120
tttaactacc	aaaaacatca	tatggggaca	agaaaaagta	ataaaacgta	tggaaaattt	180
tgtagatgtt	ataaatggat	aattattcaa	gtgataatct	atcactttga	tcttatctct	240
ttatccaatt	taattacttt	gtctctaagt	gatttgcttc	caaaatctaa	gtgtagtcta	300
tcctatttct	atcttatcct	atcatataat	cttctatata	tatgtgagtc	cgatgttgta	360
aagcgtagca	gagagagtaa	tgaagagtga	agtgttatat	tgttctctcg	tccacttcca	420
ctctctcttt	tatctctttc	ttacttcttc	gtaagatcat	tacatataat	aaataatatt	480
atthtggttt	gtgtttatatt	taataacagt	aaaaagtttt	aaaacgttga	aaaaattagc	540
cgacatagaa	tacaaaagag	gggttagcatc	gggggagaaa	cgtggaccaa	catgatacac	600
cctccaaaat	agtcaccaag	ttgaaacatt	gacatgtttc	gctttttctt	ttctgtgtat	660
actttttttt	tctgtgggtc	acattattta	atatttgtat	acaagcagct	atthttacatg	720
gagatttctt	gtcggtagat	cgctctcatt	tctccatcgc	ttccactttt	ttcctatact	780
aatttgatct	aattaattca	tatgtcaaaa	cattaagaaa	atgaaactcg	taattcatac	840
ttgaatttaa	tagattaatt	aaaatgctat	ttattggcaa	aataaactcg	gtttatatct	900
aaatthttaga	atcactaaaa	ctthttgccc	aaaaaaaaat	aaaaataaat	cactaaaaaca	960
aaaaacaatc	aaaagaaaac	ccatgttggt	aaatcggata	atgaaaaata	ttagaatccc	1020
cgtcctttgt	gtattttggc	gtagcatgaa	actatataat	aaacatgcat	tcattcttag	1080
acttctcgta	gcttatcaac	aacaacgcgc	togatctctc	tcagcctgtc	tgacaactct	1140
ttctctagtt	ctagagtttt	caatthtattg	ttgagccttt	tattaaaaaa	aaaaaaacaa	1200
gaacaaaaga	aatgggttcaa	ttgtcaagaa	aagctacatg	caacagccat	ggccaagtct	1260
cttctgtatt	ccttggttgg	gaagagtacg	agaagaatcc	ttacgacgtt	accaagaacc	1320
ctcaaggcat	tatccagatg	ggctctgcgg	aaaatcaggt	aaacaaatat	tattcaacag	1380
catgtgatat	atatatactt	atgtatatca	tgacagagag	actaatthaa	agtatgttta	1440
atthttattgg	atthctgtag	ctatgctttg	atctactaga	gtcatggctt	gcacaaaaca	1500
cagacgcagc	ctgtttcaag	agagatggcc	agtctgtttt	ccgggaactc	gctctctttc	1560
aagactacca	tggcctctct	tccttcaaaa	atgtaagatt	attaattgta	tttatcaaat	1620
ttattttagt	gttctgtatc	ttgctcgaat	gattthtcagg	cctttgctga	tttcatgtca	1680
gaaaaatagag	gaaatcgagt	ttctthttgat	tcaaacaacc	ttgtgctcac	tgctggagcc	1740
acttccgcaa	acgagactct	aatgtthttgt	cttgacagatc	ccggtgacgc	tttcttgctt	1800
cccacgccat	attatccagg	gttagtccac	tgtttgtcta	cacgtaaaaat	ttccatcatt	1860
cctacgaact	tgacttaact	aaaactcatg	tttattthttg	tacttcaggt	ttgataggga	1920
tctaaaatgg	cgaaccgggg	ttgagattgt	accaatccaa	agctcaagta	ctaaccgggtt	1980
tcgcataacg	aaacttgcac	tcgaagaagc	ctacgagcaa	gccaagaagc	ttgacctaaa	2040
cgtcaaagga	atactcatca	ccaaccatc	taaccctttg	ggtacgacaa	caacccaaac	2100
cgaactcaac	attctatttg	atthcatcac	caagaataag	aatatacatt	tagtaagtga	2160
cgagatatat	tcgggcacag	tattcaactc	ttcagaattc	atcagcgtca	tggagattct	2220
aaaaaataat	caactcgaaa	acaccgatgt	tttgaaccga	gtccacattg	tttgtagctt	2280
atctaaagat	ctaggcctcc	ctggthtttag	agttggagcc	atthactcca	atgacaaaaga	2340
tgtcatctct	gccgctacaa	aaatgtcaag	tttcggcctt	gtctcctccc	agacacaata	2400
cctactatcc	tcattattat	ctgacaagaa	gttcactaag	aactacctta	gagagaacca	2460
aaaacggctc	aagaacagac	agagaaagct	cgtgttggtg	ctagaggcca	tcgggatcaa	2520
atgtctgaag	agtaatgcgg	gactctthttg	ttgggtcgac	atgagacctc	tccttagatc	2580
taaaacgttc	gaagcggaaa	tggatctthttg	gaagaagatt	gtttacgaag	tgaagctcaa	2640
catctctcct	ggthctgctg	gccattgtga	agaaccgggt	tggthttagag	tttgtthctgc	2700
gaacatgatt	gatgagacat	taaagcttgc	tttaaaagaga	ttgaagatgt	tggthtgatga	2760

tgaaaactca	agtagaagat	gccaaaagag	taaaagcgaa	agactaaacg	gttcgaggaa	2820
gaagacgatg	tcaaatgtct	ctaactgggt	tttccgacta	tcgtttcacg	accgtgaggc	2880
tgaggaacga	tagtccggtt	tttgttttga	agttcttttt	ttttgtttcc	cacacattgc	2940
aagtgattct	gtaatttttt	ttatcacgag	agagagtgtg	aaaaaatgga	aatgcaacgt	3000
gcttactctg	atcctagatt	ttagaaaacc	gttgaagact	tcttagagca	agtccatcgg	3060
cagtttttaa	tgggtttcta	atgggtttct	agctaattaa	aagtccaaaa	ttaaatgaaa	3120
acccaactaa	ataattagga	tccatcccaa	tattaggttt	tttggatggg	tttttagacg	3180
gcgacgtggt	cgactgtgag	tcgtcggaaa	acaaaaaaaa	tcacaacact	catgttttcc	3240
tttttctctc	cgttttttcac	ttttttgttt	tgtccgacgg	ccggcgattc	gaatcgattt	3300
gatctccggt	gtatcgaaca	tgaaatcggg	agagaagagc	caaatcatcg	acgacttggt	3360
tcaccaattc	cattcttcga	accatactca	tataagagtt	tcttggtctc	tctctaaaac	3420
tottctaatt	ttctgata					3438

<210> 13
 <211> 68
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Beneficial Oligonucleotide-Contains both DNA and RNA

<400> 13	
cagggtcaagt gcaacgtagg atgatttttta ucaaccuacg ttgcacuuga ccuggcgcggt	60
tttcgcgc	68

<210> 14
 <211> 68
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Beneficial Oligonucleotide-Contains Both DNA and RNA

<400> 14	
cagggtcaagt gctacgtagg atgatttttta ucaaccuacg tagcacuuga ccuggcgcggt	60
tttcgcgc	68

<210> 15
 <211> 24
 <212> DNA
 <213> Jelly Fish

<400> 15	
atgggtgagca agggcgagga gctg	24

<210> 16
 <211> 24
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Mutation

<400> 16
 atggtgagca agggctagga gctg 24
 <210> 17
 <211> 24
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Mutation
 <400> 17
 atggtgagca agggcaggag ctgt 24
 <210> 18
 <211> 68
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Beneficial Oligonucleotide-Contains Both DNA and RNA
 <400> 18
 gtgagcaagg gcgaggagct gttcattttu gaacagcucc tcgcccucgc ucacgcgcgt 60
 tttcgcgc 68
 <210> 19
 <211> 68
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Beneficial Oligonucleotide-Contains Both DNA and RNA
 <400> 19
 tgagcaaggg ctcgagctg ttcacttttg ugaacagcuc cgagcccuug cucagcgcgt 60
 tttcgcgc 68

Pending Application 09/108,006

Filed on 06/30/98

Inventors: Steer *et al.*

Our file: 7991-015-999

ATTORNEY DOCKET NUMBER: 7991-086-99
SERIAL NUMBER: 09/685,403
REFERENCE: **BE**

HEPATOCELLULAR CHIMERAPLASTY

This application is a continuation-in-part of application Serial No. PCT/US 98/08834, filed April 30, 1998, which claims benefit of the priority of U.S. patent application Serial No. 60/045,288, filed April 30, 1997, now abandoned, and application Serial No. 60/054,837, filed August 5, 1997, and application Serial No. 60/064,996, filed November 10, 1997, each of which are hereby incorporated by reference in their entirety. This application also claims benefit of the priority of application Serial No. 60/074,497, filed February 12, 1998.

1. FIELD OF THE INVENTION

The invention concerns methods and compositions for the use of recombinagenic oligonucleobases *in vivo* for the correction of disease causing genetic defects and the prevention of disease by introducing genetic modifications into the genes that encode Apolipoprotein B (Apo B) and Apolipoprotein E (Apo E)

2. BACKGROUND TO THE INVENTION

2.1 THE USE OF CHIMERIC MUTATIONAL VECTORS TO EFFECT GENETIC CHANGES IN CULTURED CELLS

The inclusion of a publication or patent application in this specification is not an admission that the publication or the invention, if any, of the application occurred prior to the present invention or resulted from the conception of a person other than the present inventors.

The published examples of recombinagenic oligonucleobases are termed Chimeric Mutational Vectors (CMV) or chimeraplasts because they contain both 2'-O-modified ribonucleotides and deoxyribonucleotides.

An oligonucleotide having complementary deoxyribonucleotides and ribonucleotides and containing a sequence homologous to a fragment of the bacteriophage M13mp19, was described in Kmiec, E.B., et al., November 1994, Mol. and Cell. Biol. **14**, 7163-7172. The oligonucleotide had a single contiguous segment of ribonucleotides. Kmiec et al. showed that the oligonucleotide was a substrate for the REC2 homologous pairing enzyme from *Ustilago maydis*.

Patent publication WO 95/15972, published June 15, 1995, and counterpart U.S. Patent No. 5,565,350 (the '350 patent) described duplex CMV for the introduction of genetic changes in eukaryotic cells. Examples in a *Ustilago maydis* gene and in the murine ras gene were reported. The latter example was designed to introduce a transforming mutation into the ras gene so that the successful mutation of the ras gene in NIH 3T3 cells would cause the growth in soft agar of a colony of cells ("transformation"). The '350 patent reported that the maximum rate of transformation of NIH 3T3 was less than 0.1 %, i.e., about 100 transformants per 10^6 cells exposed to the ras duplex CMV. In the *Ustilago maydis* system the rate of transformants was about 600 per 10^6 . A chimeric vector designed to introduce a mutation into a human bcl-2 gene was described in Kmiec, E.B., February 1996, Seminars in Oncology 23, 188.

A duplex CMV designed to repair the mutation in codon 12 of K-ras was described in Kmiec, E.B., December 1995, Advanced Drug Delivery Reviews 17, 333-40. The duplex CMV was tested in Capan 2, a cell line derived from a human pancreatic adenocarcinoma, using LIPOFECTIN™ to introduce the duplex CMV into the Capan 2 cells. Twenty four hours after the duplex CMV was introduced, the cells were harvested and genomic DNA was extracted; a fragment containing codon 12 of K-ras was amplified by PCR and the rate of conversion estimated by hybridization with allele specific probes. The rate of repair was reported to be approximately 18%.

A duplex CMV designed to repair a mutation in the gene encoding liver/bone/kidney type alkaline phosphatase was reported in Yoon, K., et al., March 1996, Proc. Natl. Acad. Sci. 93, 2071. The alkaline phosphatase gene was transiently introduced into CHO cells by a plasmid. Six hours later the duplex CMV was introduced. The plasmid was recovered at 24 hours after introduction of the duplex CMV and analyzed. The results showed that approximately 30 to 38% of the alkaline phosphatase genes were repaired by the duplex CMV.

WO 97/41411 and counterpart United States Patent No. 5,760,012 to E.B. Kmiec, A. Cole-Strauss and K. Yoon, and the publication Cole-Strauss, A., et al., September 1996, SCIENCE 273, 1386 disclose duplex CMV that are used in the treatment of genetic diseases of hematopoietic cells, e.g., Sickle Cell Disease, Thalassemia and Gaucher Disease. United States Patent Application Serial No. 08/664,487, filed June 17,

1996, by E.B. Kmiec describes duplex CMV having non-natural nucleotides for use in specific, site-directed mutagenesis. The duplex CMV described in the applications and certain of the publications of Kmiec and his colleagues contain a central segment of DNA:DNA homoduplex and flanking segments of RNA:DNA hybrid-duplex or 2'-OMe-RNA:DNA hybrid-duplex.

The work of Kmiec and his colleagues concerned cells that are mitotically active, i.e., proliferating cells, at the time they are exposed to CMV. Kmiec and colleagues used a CMV/liposomal macromolecular carrier complex in which the CMV were mixed with a pre-formed liposome or lipid vesicle. In such a complex the CMV are believed to adhere to the surface of the liposome.

Kren et al., June 1997, *Hepatology* 25, 1462-1468, reported the successful use of a CMV in non-replicating, primary tissue-cultured rat hepatocytes to mutate the coagulation factor IX gene. Kren et al., March 1998, *Nature Medicine* 4, 285 reported the use of a CMV *in vivo* to introduce a genetic defect in the same gene.

2.2 THE USE OF A POLYETHYLENIMINE MACROMOLECULAR CARRIER FOR *IN VIVO* AND *IN VITRO* TRANSFECTION

Branched chain polyethylenimine has been used as a carrier to introduce nucleic acids into eukaryotic cells both *in vivo* and *in vitro*. Boussif, O., et al., 1995, *Proc. Natl. Acad. Sci.* 92, 7297; Abdallah, B. et al., 1996, *Human Gene Therapy* 7, 1947. Boletta, A., et al., 1997, 8, 1243-1251. The *in vitro* use of galactosylated polyethylenimine to introduce DNA into cultured HepG2 hepatocarcinoma cell lines is reported by Zanta, et al., October 1, 1997, *Bioconjugate Chemistry* 8, 839-844. The coupling of a protein ligand, transferrin, to polyethylenimine and its use to introduce a test gene into cultured cells by use of the transferrin receptor is described in Kircheis, R., et al., 1997, *Gene Therapy* 4, 409-4-18. Branched chain polyethylenimines contain secondary and tertiary amino groups having a broad range of pK's and, consequently these polyethylenimines have a substantial buffering capacity at a pH where polylysine has little or no capacity, i.e., less than about 8. Tang, M.K., & Szoka, F.C., 1997, *Gene Therapy* 4, 823-832. The use of branched chain polyalkanylimines, including polyethylenimine as carriers for the introduction of nucleic acids into cells is described in WO 96/02655 to J-P. Behr et al.

The successful *in vivo* and *in vitro* use of linear polyethylenimine to transfect a gene is reported by Ferrari, S., et al., 1997, *Gene Therapy* **4**, 1100-1106. Compositions comprising a linear polyalkanylimine and a nucleic acid as disclosed in patent publication WO 93/20090 to S. Stein et al.

2.3 THE USE OF A LIPOSOMAL CARRIER FOR *IN VIVO* TRANSFECTION

The use of liposomes or lipid vesicles to introduce DNA encoding a foreign protein into cells has been described. The most frequently used techniques adhere the DNA to the surface of a positively charged liposome, rather than encapsulating the DNA, although encapsulated DNA techniques were known. United States Patent Nos. 4,235,871 and 4,394,448 are relevant. The field is reviewed by Smith, J.G., et al., 1993, *Biochim. Biophys. Acta* **1154**, 327-340 and Staubinger, R.M., et al., 1987, *Methods in Enzymology* **185**, 512. The use of DOTAP, a cationic lipid in a liposome to transfect hepatic cells *in vivo* is described in Fabrega, A.J., et al., 1996, *Transplantation* **62**, 1866-1871. The use of cationic lipid-containing liposomes to transfect a variety of cells of adult mice is described in Zhu, N., et al., 1993, *Science* **261**, 209. The use of phosphatidylserine containing lipids to form DNA encapsulating liposomes for transfection is described in Fraley, R., et al., 1981, *Biochemistry* **20**, 6978-87.

2.4 THE USE OF THE ASIALOGLYCOPROTEIN RECEPTOR FOR HEPATOCELLULAR SPECIFIC TRANSFECTION

United States Patent Nos. 5,166,320 and 5,635,383 disclose the transfection of hepatocytes by forming a complex of a DNA, a polycationic macromolecular carrier and a ligand for the asialoglycoprotein receptor. In one embodiment, the macromolecular carrier was polylysine. The use of a lactosylcerebroside containing liposome to transfect a hepatocyte *in vivo* is described by Nandi, P.K., et al., 1986, *J. Biol. Chem.* **261**, 16722-16722. The use of asialofetuin-labeled liposomes to transfect liver cells with a reporter plasmid is described in Hara et al., 1995, *Gene Therapy* **2**, 764-788. The use of galactosylated polyethylenimine to transfect cultured hepatocytes is described in Zanta M-A., et al. abst. pub. Oct. 1, 1997, *Bioconjugate Chem.*, **8**, 839-844.

2.5 APO B100, APO B48 AND THE REDUCTION OF SERUM LDL

Hepatic and Intestinal Lipoprotein Secretion: Both the liver and the intestines make and export lipoproteins for the transport of lipids. The lipoproteins are termed very low density lipoproteins (VLDL) and chylomicrons, respectively. VLDL and chylomicrons differ in size and in their major protein components. The major protein of VLDL is Apo B100, consisting of 4536 amino acids; the major protein of chylomicrons is Apo B48, which consists of the N-terminal 2152 amino acids of Apo B100. Apo B48 and Apo B100 are encoded by a single gene, the transcript of which is modified at nucleotide 6666 (codon 2179) by a sequence specific cytidine deaminase, termed apolipoprotein B mRNA editing enzyme (APOBE). The action of this enzyme converts a C to U and results in a stop codon.

Both VLDL, which contain Apo B100, and chylomicrons, which contain Apo B48 transport triglycerides in the vascular system to a delivery site. However, after triglyceride hydrolysis and delivery VLDL are transformed into LDL, while chylomicrons are not. High levels of circulating LDL *per se* and a high LDL:HDL ratio increase the risk of arterial atherosclerosis. Hence, it has been suggested that increasing the ratio of Apo B48 to Apo B100 would have a beneficial effect.

In many species of mammals, e.g., rats and mice, a high percentage of the lipid secretions of both liver and intestine contain Apo B48. Such species have markedly lower ratios of LDL:HDL. Greve J., et al., 1995, Proc. Zool. Soc., Calcutta, **47**, 93-100. In others, such as humans and rabbits, hepatocytes lack APOBE and the hepatocytes consequently produce only VLDL.

One strategy to reduce the atherosclerosis in humans has been to introduce the gene for the catalytic component of the apolipoprotein B editing enzyme (APOBEC-1) under the control of a constitutive promoter to convert Apo B100 transcripts into Apo B48 transcripts. The transient expression of APOBEC-1 in the hepatocytes of normal and genetically hyperlipidemic Watanabe rabbit does cause a transient reduction in the levels of LDL. Greeve, J., et al., 1996, J. Lipid Res. **37**, 2001-17. However, the uncontrolled production of APOBEC-1 is mutagenic and may cause hepatocellular hyperplasia and hepatocellular carcinoma. Yamanaka, S., et al., 1995, Proc. Natl. Acad. Sci. **92**, 8483-8487.

Individuals who are homozygous or mixed heterozygotes for genes encoding truncated Apo B100 have been observed. Malloy et al., 1981, J. Clin. Invest. 67, 1441; Hardman, D.A., et al., 1991, J. Clin. Invest. 88, 1722. These individuals have low or absent LDL. For example, deletion of nucleotides 5391-5394 results in a frame shift mutation and a shortened Apo B (B37). These patients are most often asymptomatic. Steinberg, D., et al., 1979, J. Clin. Invest. 64, 292; Young, S.G., et al., 1988, Science 241, 591; Young, S.G., 1987, J. Clin. Invest. 79, 1831. Reviewed Linton, M.F., 1993, J. Lipid. Res. 34, 521; Kane, J.P. & Havel, R.J., 1995, Chapt. 57, THE METABOLIC BASIS OF INHERITED DISEASE, ed. Scriver et al. (McGraw Hill, New York). Similarly, as many as 1 in every 3,000 persons has a serum cholesterol level of 100 mg/dl or less because the individual is heterozygous for a truncated Apo B gene. *Ibid.*, p. 1866.

Truncations that result in an Apo B that are shorter than Apo B 31 do not circulate. Truncated Apo B 86, 87 and 90 have been observed. Apo B 86 and Apo B 87, are not associated with LDL while Apo B 90 is. Each mutation is associated with hypobetalipoproteinemia. Linton, M.L., et al., 1990, Clin. Res. 38, 286A (abstr.); Tennyson, G.E., et al., 1990, Clin. Res. 38, 482A (abstr.); Kruhl, E.S., et al., 1989, Arteriosclerosis 9, 856.

2.6 APO E POLYMORPHISM AND TYPE III HYPERLIPIDEMIA

Apolipoprotein E is the major ligand for the LDL receptor for lipoproteins that contain Apo B48. There are three allelic forms of human Apo E that differ from each other by one or two amino acids: Apo E2 (Cys¹¹² Cys¹⁵⁸); Apo E3 (Cys¹¹² Arg¹⁵⁸); and Apo E4 (Arg¹¹² Arg¹⁵⁸). There is considerable geographical variation in the prevalences of the alleles. Excluding Africa, E2 ranges between 4% and 12 %, E3 between 70% and 85% and E4 between 7.5 and 25%. In the Sudan, the prevalences are 8.1%, 61.9% and 29.1%, respectively. Mahley, R.W. & Rall, S.C., Jr., 1995, Chapt. 61, THE METABOLIC BASIS OF INHERITED DISEASE, ed. Scriver et al. (McGraw Hill, New York). Thus approximately 1% of the North American and European population are Apo E 2/2 homozygotes. Of these homozygotes approximately between 2% and 10% display type III hyperlipidemia. Paradoxically, however, Apo E 2/2 homozygotes that have not

developed overt Type III hyperlipidemia display lower than average LDL associated cholesterol. Davignon, J., 1988, *Arteriosclerosis* **8**, 1.

The E4 allele is also associated with increased incidence of a major disease, Alzheimer's Disease, and with increased risk of coronary artery disease. Roses, A.D., 1996, *Ann. NY Acad. Sci.* **802**, 50-57; Okumoto, K., & Fujiki, Y., 1997, *Nature Genetics* **17**, 263; Kuusi, T., et al., 1989, *Arteriosclerosis* **9**, 237. A polymorphism in the region 491 nt 5' to the transcription start site of the Apo E gene is also and independently associated with increased risk of Alzheimer's disease. Individuals homozygous for the -491-A genotype have an increased risk of Alzheimer's, while individuals homozygous or heterozygous for the -491 T genotype have no increased risk. Bullido, M.J., 1998, et al., *Nature Genetics* **18**, 69-71.

The E2 allele in most individuals is associated with the lowest levels of serum cholesterol and LDL. However, about 5% of E2/E2 homozygous persons who are subject to environmental or genetic stress develop type III hyperlipidemia. The most common stressors are hypothyroidism, untreated diabetes mellitus, alcoholism and marked weight gain. Removal of the stressor usually results in control of the hyperlipidemia. Rare patients with type III hyperlipidemia have mutant Apo E genes. Mahley & Rall, *ibid.* Table 61-5.

3. SUMMARY OF THE INVENTION

The present invention concerns methods of treatment and/or prophylaxis which consists of the introduction of specific genetic alterations in genes of a subject individual. In one embodiment, the specific genetic alteration blocks the synthesis of Apo B100 and thereby reduces the level of LDL cholesterol. In an alternative embodiment, the specific alteration converts an Apo E4 allele to an Apo E3 or Apo E2 allele, which is associated with decreased risk of atherosclerosis and Alzheimer's Disease. In further alternative embodiments, the invention concerns the correction of inherited genetic defects in the genes of hepatocytes of individuals having a disease caused by such defects.

The invention can be practiced using any oligonucleotide or analog or derivative thereof, now known or hereafter developed, that can cause specific genetic alterations in

the genome of the hepatocytes of the subject individual (hereafter a "recombinagenic oligonucleobase"), for example a chimeric mutational vector (CMV) as, for example, described in United States patent No. 5,565,350, No. 5,731,181, and No. 5,760,012. Alternatively, the recombinagenic oligonucleobase can be a heteroduplex mutational vector or a non-chimeric mutational vector as described in U.S. patent application No. 09/078,063 and No. 09/078,064, filed May 12, 1998, each of which are hereby incorporated by reference.

In a preferred embodiment the recombinagenic oligonucleobase is complexed with a macromolecular carrier to which is attached a specific ligand. The ligand is selected to bind to a cell-surface receptor that is internalized into hepatocytes through clathrin-coated pits into endosomes. The cell surface receptors that bind such ligands are termed herein "clathrin-coated pit receptors". Examples of hepatic clathrin-coated pit receptors include the low density lipoprotein (LDL) receptor and the asialoglycoprotein receptor.

In specific embodiments the macromolecular carrier can be 1) an aqueous-cored lipid vesicle of between 25 nm and 400 nm diameter, wherein the aqueous core contains the CMV; 2) a lipid nanosphere of between 25 nm and 400 nm diameter, having a lipid core, wherein the lipid core contains a lipophilic salt of the CMV; or 3) a polycationic salt of the CMV. Examples of polycations for such salts include polyethylenimine, polylysine and histone H1. In one embodiment the polycation is a linear polyethylenimine (PEI) salt having a mass average molecular weight greater than 500 daltons and less than 1.3 Md. Alternatively the polycation can be a branched-chain polyethylenimine.

4. BRIEF DESCRIPTION OF THE FIGURE

Figure 1 is a schematic of one embodiment of CMV useful in the invention.

Figures 2A-2C show the genomic sequence of human APO E gene with translation of exons. Introns are in lower case and exons are in upper case.

60, 102, 162

5. DEFINITIONS

The invention is to be understood in accordance with the following definitions.

An oligonucleobase is a polymer of nucleobases, which polymer can hybridize by Watson-Crick base pairing to a DNA having the complementary sequence.

Nucleobases comprise a base, which is a purine, pyrimidine, or a derivative or analog thereof. Nucleobases include peptide nucleobases, the subunits of peptide nucleic acids, and morpholine nucleobases as well as nucleosides and nucleotides.

Nucleosides are nucleobases that contain a pentosefuranosyl moiety, e.g., an optionally substituted riboside or 2'-deoxyriboside. Nucleosides can be linked by one of several linkage moieties, which may or may not contain a phosphorus. Nucleosides that are linked by unsubstituted phosphodiester linkages are termed nucleotides.

An oligonucleobase chain has a single 5' and 3' terminus, which are the ultimate nucleobases of the polymer. A particular oligonucleobase chain can contain nucleobases of all types. An oligonucleobase compound is a compound comprising one or more oligonucleobase chains that are complementary and hybridized by Watson-Crick base pairing. Nucleobases are either deoxyribo-type or ribo-type. Ribo-type nucleobases are pentosefuranosyl containing nucleobases wherein the 2' carbon is a methylene substituted with a hydroxyl, alkyloxy or halogen. Deoxyribo-type nucleobases are nucleobases other than ribo-type nucleobases and include all nucleobases that do not contain a pentosefuranosyl moiety.

An oligonucleobase strand generically includes both oligonucleobase chains and segments or regions of oligonucleobase chains. An oligonucleobase strand has a 3' end and a 5' end. When a oligonucleobase strand is coextensive with a chain, the 3' and 5' ends of the strand are also 3' and 5' termini of the chain.

A region is a portion of an oligonucleobase, the sequence of which is derived from some particular source, e.g., a CMV having a region of at least 15 nucleotides having the sequence of a fragment of the human β -globin gene. A segment is a portion of a CMV having some characteristic structural feature. A given segment or a given region can contain both 2'-deoxynucleotides and ribonucleotides. However, a ribo-type segment or a 2'-deoxyribo-type segment contain only ribo-type and 2'-deoxyribo-type nucleobases, respectively.

6. DETAILED DESCRIPTION OF THE INVENTION

6.1 THE STRUCTURE OF THE CHIMERIC MUTATIONAL VECTOR

The Chimeric Mutational Vectors (CMV) are comprised of oligonucleobases, i.e., polymers of nucleobases, which polymers form Watson-Crick base pairs of purines and pyrimidines (hybridize), to DNA having the appropriate sequence. Each CMV is divided into a first and a second strand of at least 15 nucleobases each that are complementary to each other. The strands can be, but need not be, covalently linked. Nucleobases contain a base, which is either a purine or a pyrimidine or analog or derivative thereof. There are two types of nucleobases. Ribo-type nucleobases are ribonucleosides having a 2'-hydroxyl, substituted 2'-hydroxyl or 2'-halo-substituted ribose. All nucleobases other than ribo-type nucleobases are deoxyribo-type nucleobases. Thus, deoxy-type nucleobases include peptide nucleobases. As used herein, only a recombinagenic oligonucleobase that contains at least three contiguous ribo-type nucleobases that are hybridized to deoxyribo-type nucleobases are considered CMV.

The sequence of the first and second strands consists of at least two regions that are homologous to the target gene, i.e., have the same sequence as fragments of the target gene, and one or more regions (the "mutator regions") that differ from the target gene and introduce the genetic change into the target gene. The mutator region is located between homologous regions. In certain embodiments of the invention, each of the flanking homologous regions contains a ribo-type segment of at least three ribo-type nucleobases, that form a hybrid duplex, preferably at least six ribo-type nucleobases and more preferably at least ten ribo-type nucleobases in length, but not more than 25 and preferably not more than 20, more preferably not more than 15 ribo-type nucleobases. The hybrid-duplex-forming ribo-type oligonucleobase segments need not be adjacent to the mutator region. In certain embodiments of the invention the ribo-type oligonucleobase segments are separated from the mutator region by a portion of the homologous region comprising deoxyribo-type nucleobases. In these embodiments the mutator region is also composed of deoxyribo-type nucleobases. Accordingly, the mutator region and a portion of one or both homologous regions form an intervening segment of homo-duplex, which separates the two segments of hybrid-duplex.

The total length of all homologous regions is preferably at least 16 nucleobases and is more preferably from about 20 nucleobases to about 60 nucleobases in length.

Preferably, the mutator region consists of 20 or fewer bases, more preferably 6 or fewer bases and most preferably 3 or fewer bases. The mutator region can be of a length different than the length of the sequence that separates the regions of the target gene homology with the homologous regions of the CMV so that an insertion or deletion of the target gene results. When the CMV is used to introduce a deletion in the target gene there is no base identifiable as within the mutator region. Rather, the mutation is effected by the juxtaposition of the two homologous regions that are separated in the target gene. For the purposes of the invention, the length of the mutator region of a CMV that introduces a deletion in the target gene is deemed to be the length of the deletion. In one embodiment the mutator region is a deletion of from 6 to 1 bases or more preferably from 3 to 1 bases. Multiple separated mutations can be introduced by a single CMV, in which case there are multiple mutator regions in the same CMV. Alternatively multiple CMV can be used simultaneously to introduce multiple genetic changes in a single gene or, alternatively to introduce genetic changes in multiple genes of the same cell. Herein the mutator region is also termed the heterologous region.

In one embodiment the CMV is a single oligonucleobase chain of between 40 and 100 nucleobases. In an alternative embodiment, the CMV comprises a first and a second oligonucleobase chain, each of between 20 and 100 bases; wherein the first chain comprises the first strand and the second chain comprises the second strand. The first and second chains can be linked covalently by other than nucleobases or, alternatively, can be associated only by Watson-Crick base pairings. In an alternative embodiment the CMV is a first strand which is a single oligonucleobase chain and a second strand, complementary to the first which consists of two oligonucleobase chains, which are linked to the first strand chain by linkers. The combined length of the two chains of the second strand is the length of the first strand.

Linkers: Covalent linkage of the first and second strands can be made by oligo-alkanediols such as polyethyleneglycol, poly-1,3-propanediol or poly-1,4-butanediol. The length of various linkers suitable for connecting two hybridized nucleic acid strands is understood by those skilled in the art. A polyethylene glycol linker having from six to

three ethylene units and terminal phosphoryl moieties is suitable. Durand, M. et al., 1990, *Nucleic Acid Research* **18**, 6353; Ma, M. Y-X., et al., 1993, *Nucleic Acids Res.* **21**, 2585-2589. A preferred alternative linker is bis-phosphorylpropyl-trans-4,4'-stilbenedicarboxamide. Letsinger, R.L., et alia, 1994, *J. Am. Chem. Soc.* **116**, 811-812; Letsinger, R.L. et alia, 1995, *J. Am. Chem. Soc.* **117**, 7323-7328, which are hereby incorporated by reference. Such linkers can be inserted into the DMV using conventional solid phase synthesis. Alternatively, the strands of the DMV can be separately synthesized and then hybridized and the interstrand linkage formed using a thiophoryl-containing stilbenedicarboxamide as described in patent publication WO 97/05284, February 13, 1997, to Letsinger R.L. et alia.

In a further alternative embodiment the linker can be a single strand oligonucleobase comprised of nuclease resistant nucleobases, e.g., a 2'-O-methyl, 2'-O-allyl or 2'-F-ribonucleotides. The tetranucleotide sequences TTTT, UUUU and UUCG and the trinucleotide sequences TTT, UUU, or UCG are particularly preferred nucleotide linkers.

Nucleotides: In an alternative embodiment the invention can be practiced using CMV comprising deoxynucleotides or deoxynucleosides and 2'-O substituted ribonucleotides or ribonucleosides. Suitable substituents include the substituents taught by the Kmiec Application, C₁₋₆ alkane. Alternative substituents include the substituents taught by U.S. Patent No. 5,334,711 (Sproat) and the substituents taught by patent publications EP 629 387 and EP 679657 (collectively, the Martin Applications), which are hereby incorporated by reference. As used herein a 2' fluoro, chloro or bromo derivative of a ribonucleotide or a ribonucleotide having a substituted 2'-O as described in the Martin Applications or Sproat is termed a "2'-Substituted Ribonucleotide." Particular preferred embodiments of 2'-Substituted Ribonucleotides are 2'-fluoro, 2'-methoxy, 2'-propyloxy, 2'-allyloxy, 2'-hydroxyethyloxy, 2'-methoxyethyloxy, 2'-fluoropropyloxy and 2'-trifluoropropyloxy substituted ribonucleotides. In more preferred embodiments the 2'-Substituted Ribonucleotides are 2'-fluoro, 2'-methoxy, 2'-methoxyethyloxy, and 2'-allyloxy substituted nucleotides.

2'-Substituted Ribonucleosides are defined analogously. Particular preferred embodiments of 2'-Substituted Ribonucleosides are 2'-fluoro, 2'-methoxy, 2'-propyloxy,

2'-allyloxy, 2'-hydroxyethyloxy, 2'-methoxyethyloxy, 2'-fluoropropoxy and 2'-trifluoropropoxy substituted ribonucleotides. In more preferred embodiment on the 2'-Substituted Ribonucleosides are 2'-fluoro, 2'-methoxy, 2'-methoxyethyloxy, and 2'-allyloxy substituted nucleotides.

The term "nuclease resistant ribonucleoside" encompasses 2'-Substituted Ribonucleosides, including 2'-Substituted Ribonucleotides and also all 2'-hydroxyl ribonucleosides other than ribonucleotides. In a preferred embodiment, the CMV preferably includes at least three and more preferably six nuclease resistant ribonucleosides. In one preferred embodiment the CMV contains no nuclease sensitive ribonucleosides. In an alternative preferred embodiment, every other ribonucleoside is nuclease resistant. Certain 2'-blocking groups can be more readily synthesized for purines or pyrimidines. In one embodiment of the CMV only the ribonucleoside purines or only the ribonucleoside pyrimidines are nuclease resistant.

Recombinagenic oligonucleobases, including non-chimeric mutational oligonucleobases and improved CMV and their use in eukaryotic cells and cell-free systems are described in U.S. patent applications Serial No. 09/078,063, filed May 12, 1998, and Serial No. 09/078,064, filed May 12, 1998, which are each hereby incorporated in their entirety. These mutational oligonucleobases can be used in the same manner as the CMV described in this application.

6.2 THE GENE-SPECIFIC STRUCTURE OF THE CHIMERIC MUTATIONAL VECTOR

Figure 1 shows a diagram of a CMV according to one embodiment of the invention. In the Figure segments "a" and "c-e" are target gene specific segments of the CMV. The sequence of segment "a" and "c-e" are complements of each other. The sequence of segments "f" and "h" are also complements of each other but are unrelated to the specific target gene and are selected merely to ensure the stability of hybridization in order to protect the 3' and 5' ends. Additional protection of the 3' and 5' ends can be accomplished by making the 5' and 3' most internucleotide bonds a phosphorothioate, phosphonate or any other nuclease resistant bond. The sequence of segments "f" and "h" can be 5'-GCGCG-3' or permutations thereof. Segments "g" and "b" can be any linker that covalently connects the two strands, e.g., four unpaired nucleotides or an alkoxy

oligomer such as polyethylene glycol. When segments "g" and "b" are composed of other than nucleobases, then segments "a", "c-f" and "h" are each an oligonucleobase chain.

The ribo-type nucleobase segments are segments "c" and "e," which form hybrid-duplexes by Watson-Crick base pairing to the complementary portions of segment "a." The segment "a" can have the sequence of either the coding or non-coding strand of the gene.

Table I contains SEQ ID No. 4 - No. 24, which are examples of the sequences that can be used to practice the invention. The mutator region in each case is underlined and in bold. CMV having a segment "a" with a sequence selected from the sequences of Table I can be used to practice the invention. Alternatively, segment "a" may have the sequence of the complement of a sequence of Table I. As used herein, a CMV or other type of recombinagenic oligonucleobase comprises a sequence if either strand of the CMV or recombinagenic oligonucleobase comprises the sequence or comprises a sequence containing ribo-type nucleobases with uracil bases replacing thymine bases. Thus, for example, a CMV having the sequence 5'-agucuggaugGGTAAgccgcccuca-3' (SEQ ID No. 26) is considered to have the sequence of SEQ ID No: 4, wherein the lower case letters denote ribo-type nucleobases and the UPPER CASE letters denote deoxyribo-type nucleobases.

Subjects can be treated with a recombinagenic oligonucleobase specific for Apo B or Apo E according to the guidance of the Factor IX example below. More particularly the recombinagenic oligonucleobase can be given in divided doses at intervals that permit determining of the phenotypic effect of the dose, i.e., evaluation of the extent of the decline in LDL cholesterol and observation for adverse reactions. A reduction of the subject's fasting LDL serum cholesterol to below the level of the 5th percentile of the age-matched population (80-90 mg/dl) can be used as a therapeutic end point; alternatively reduction of fasting LDL serum cholesterol to below the average age-matched normal value (100-140) can be used. The number and size of the dose(s) can be modified to control the extent of the phenotypic effects. In the event that reversal of the specific genetic changes appear desirable, a recombinagenic oligonucleobase having a sequence appropriate to reverse of the specific changes can be administered so that the fraction of unmodified Apo B or Apo E genes can be increased. Modification of the dose size and

number and the administration of a reversing recombinagenic oligonucleobase permits the adjustment of the number of altered genes in the subject so that a predetermined amount of the phenotypic change can be effected.

6.2.1 Specific Alterations of the Apo B Gene

SEQ ID No. 1 contains the Apo B amino acid sequence and SEQ ID No. 2 contains the Apo B cDNA sequence.

The level of serum cholesterol and particularly of LDL-associated cholesterol can be reduced in a subject by introducing mutations into the subject's hepatic Apo B genes. The mutation can be any mutation that causes termination of the Apo B translation product between amino acid 1433 (Apo B 31) and amino acid 3974 (Apo B 87). (The amino acid numbering for Apo B in this specification refers to the 4553 amino acid primary translation product, i.e., mature Apo B100 plus the 27 amino acid leader sequence. Mature Apo B 100 consists of 4536 amino acids and mature Apo B 48 consists of 2152 amino acids.) Preferably the translation product is terminated between amino acids 1841 (Apo B 40) and 2975 (Apo 65). The translation product can be terminated by introducing a frameshift mutation, i.e., by adding or deleting one or two nucleotides from the gene, or by introducing a stop codon (a TAA, TAG or TGA). The preferred stop codon is TAA. To monitor the introduction of the mutation it is preferred to have the mutation introduce or remove a palindromic sequence, which is the substrate of a restriction enzyme.

The sequence of the CMV is selected to have two homologous regions of at least 10 nucleobases and preferably at least 12 nucleobases each with a fragment of the Apo B gene located between nucleotides encoding amino acid 1433 (nt 4425) and 3974 (nt 12,048) and preferably located between the nucleotides encoding amino acids 1841 (nt 5649) and 2975 (nt 9051). In this specification, nt 6666 is the first nucleotide of codon 2180, the nucleotide that is converted by APOBE. In a preferred embodiment, the two homologous regions are separated by a single nucleobase in the sequence of the Apo B gene, where the CMV introduces a base substitution in the Apo B gene. Alternatively, the two homology regions can be adjacent in the Apo B gene and separated by a single or double nucleobase in the CMV, such that a one or two base insertion results from the

action of the CMV on the Apo B gene. Alternatively, the homologous regions can be separated in the Apo B gene by one or two nucleotides that are deleted from the sequence of the CMV, such that the action of the CMV results in a one or two base deletion in the gene.

Nucleotides 4425-12,048 of the Apo B cDNA are encoded by exon 26 (nt 4342-11913), exon 27 (nt 11914 - 12028) and exon 28 (nt 12029-12212); see Table I, and GENBANK Accession No. 19828, which is hereby incorporated by reference. When an alteration is to be made at a position 3' of nt 11913, attention must be paid to the exon/intron boundary. Mutations that are located within 10-15 nucleotides of the exon/intron boundary must be identified so that the homology region of the CMV continues with the sequence of the intron and not the exon.

The homologous regions can be each from 10 to about 15 nucleobases in length; the two regions need not be of the same length. The fraction of nucleobases that contain a guanine or cytosine base is a design consideration (the GC fraction). It is preferred that when the homologous region contains 12 or fewer nucleobases, the GC fraction be at least 33% and preferably at least 50%. When the GC fraction is less than 33% the length of the homologous regions is preferably 13, 14 or 15 nucleobases.

Table I contains 17 exemplary embodiments, SEQ ID No. 4-20, of CMV sufficient for the practice of the embodiments of the invention described in this section. Suitable CMV can be made using nt 3-23 of SEQ ID No. 4-10, 12, and 16-20. SEQ ID NO. 11 and 13-15 have a lower GC fraction; CMV sufficient for the practice of the invention can be made containing residues 3-25 of SEQ ID NO. 11 and 13-15.

6.2.2 Specific Alterations of the Apo E Gene

In a further embodiment, the invention consists of introducing specific alterations to the Apo E gene. E4 homozygous individuals are at increased risk for atherosclerosis, particularly coronary artery disease, and Alzheimer's disease. Therefore, one embodiment of the present invention is the introduction of the substitution Arg→Cys at residues 112, to convert an E4 allele to an E3 allele, and optionally at residue 158 to convert an E3 or E4 allele into an E2 allele of an Apo E gene of an hepatocyte of a subject. The substitutions can be introduced using an oligonucleobase containing the

sequence of nt 3-23 of SEQ ID No. 22 and No. 23 or complement thereof and more preferably of an oligonucleobase containing SEQ ID No. 22 and No. 23 or complement thereof. In addition, in individuals lacking genetic or environment stressors, the E2 allele results in a lowered LDL level and a decreased risk of atherosclerosis and coronary artery disease. Thus, these risks in an E3/E3 individual can be reduced by introduction of the (Arg→Cys)¹⁵⁸ substitution to convert the individual Apo E genes to E2 alleles.

Apo E2/E2 homozygous individuals who are suffering from Type III hyperlipidemia can be treated by converting E2 alleles to E3 alleles by making a Cys→Arg¹⁵⁸ substitution. Such a substitution can be made using an oligonucleobase containing the sequence of nt 3-23 of SEQ ID No. 24 or complement thereof and more preferably of an oligonucleobase containing SEQ ID No. 24 or complement thereof.

Independent of the Apo E allele, individuals who are homozygous for -491-A are at increased risk to develop Alzheimer's Disease. Bullido, M.J., 1998, et al., *Nature Genetics* **18**, 69-71. These individuals can be advantageously treated with an oligonucleobase containing the sequence of nt 3-23 of SEQ ID No. 25.

6.2.3 Repair of Mutations of the Apo B and Apo E Gene

SEQ ID No. 3 contains the Apo E genomic DNA sequence.

A further embodiment of the invention concerns the use of CMV to repair mutations in the Apo B and Apo E genes that cause hypobetalipoproteinemia and dysbetalipoproteinemia, respectively. Mutations that are located within 10-15 nucleotides of the exon/intron boundary must be identified so that the homology region of the CMV continues with the sequence of the intron and not the exon. The genomic sequence of Apo E4 indicating the exon and intron boundaries is given in Paik et al., 1985, *Proc. Natl. Acad. Sci.* **82**, 3445, which is hereby incorporated by reference. The exon/intron boundaries of the Apo B gene are given in Table II along with the GENBANK accession numbers for the genomic sequence of Apo B.

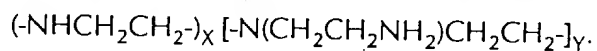
6.3 FORMULATIONS SUITABLE FOR *IN VIVO* USE

The prior art formulations of CMV and a macromolecular carrier are of limited utility for *in vivo* use because of their low capacity for CMV and because the CMV is not protected from extracellular enzymes. The invention provides three alternative macromolecular carriers that overcome the limitations of the prior art. The carriers are polyethylenimine (PEI), aqueous-cored lipid vesicles, which are also termed unilamellar liposomes and lipid nanospheres.

Each of the carriers can be further provided with a ligand that is complementary to a cell-surface protein of the target cell. Such ligands are useful to increase both the amount and specificity of the uptake of CMV into the targeted cell. In one embodiment of the invention the target cell is a hepatocyte and the ligand is a galactose saccharide or lactose disaccharide that binds to the asialoglycoprotein receptor.

6.3.1 Polycationic Carriers

The invention can be practiced using any polycation that is non-toxic when administered to cells *in vitro* or to subjects *in vivo*. Suitable examples include polybasic amino acids such as polylysine, polyarginine, basic proteins such as histone H1, and synthetic polymers such as the branched-chain polyethylenimine:



The invention can be practiced with any branched chain polyethylenimine (PEI) having an average molecular weight of greater than about 500 daltons, preferably greater than between about 10 Kd and more preferably about 25 Kd (mass average molecular weight determined by light scattering). The upper limit of suitability is determined by the toxicity and solubility of the PEI. Toxicity and insolubility of molecular weights greater than about 1.3 Md makes such PEI material less suitable. The use of high molecular weight PEI as a carrier to transfect a cell with DNA is described in Boussif, O. et al., 1995, Proc. Natl. Acad. Sci. **92**, 7297, which is hereby incorporated by reference. PEI solutions can be prepared according to the procedure of Boussif et al.

The CMV carrier complex is formed by mixing an aqueous solution of CMV and a neutral aqueous solution of PEI at a ratio of between 9 and 4 PEI nitrogens per CMV phosphate. In a preferred embodiment the ratio is 6. The complex can be formed, for

example, by mixing a 10 mM solution of PEI, at pH 7.0 in 0.15 M NaCl with CMV to form a final CMV concentration of between 100 and 500 nM.

In addition a ligand for a clathrin-coated pit receptor can be attached to the polycation or to a fraction of the polycations. In one embodiment the ligand is a saccharide or disaccharide that binds to the asialoglycoprotein receptor, such as lactose, galactose, or N-acetylgalactosamine. Any technique can be used to attach the ligands. The optimal ratio of ligand to polyethylene subunit can be determined by fluorescently labeling the CMV and injecting fluorescent CMV/molecular carrier/ligand complexes directly into the tissue of interest and determining the extent of fluorescent uptake according to the method of Kren et al., 1997, *Hepatology* **25**, 1462-1468.

Good results can be obtained using a 1:1 mixture of lactosylated PEI having a ratio of 0.4-0.8 lactosyl moieties per nitrogen and unmodified PEI. The mixture is used in a ratio of between 4 and 9 PEI nitrogens per CMV phosphate. A preferred ratio of oligonucleotide phosphate to nitrogen is 1:6. Good results can be obtained with PEIs having a mass average molecular weight of 25 Kd and 800 Kd which are commercially available from Aldrich Chemical Co., Catalog No. 40,872-7 and 18,197-8, respectively. Linear PEI such as that described in Ferrarri, S., et al., 1997, *Gene Therapy* **4**, 1100-1106 and sold under the trademenark EXGEN 500™ is particularly suitable for the practice of the invention because of its lower toxicity compared to branched-chain PEI.

In an alternative embodiment the polycationic carrier can be a basic protein such as histone H1, which can be substituted with a ligand for a clathrin-coated pit receptor. A 1:1 (w/w) mixture of histone and CMV can be used to practice the invention.

6.3.2 Lipids that Are Useful in Carriers

The selection of lipids for incorporation into the lipid vesicle and lipid nanosphere carriers of the invention is not critical. Lipid nanospheres can be constructed using semi-purified lipid biological preparations, e.g., soybean oil (Sigma Chem. Co.) and egg phosphatidyl choline (EPC) (Avanti Polar Lipids). Other lipids that are useful in the preparation of lipid nanospheres and/or lipid vesicles include neutral lipids, e.g., dioleoyl phosphatidylcholine (DOPC), and dioleoyl phosphatidyl ethanolamine (DOPE), anionic lipids, e.g., dioleoyl phosphatidyl serine (DOPS) and cationic lipids, e.g., dioleoyl

trimethyl ammonium propane (DOTAP), dioctadecyldiamidoglycyl spermine (DOGS), dioleoyl trimethyl ammonium (DOTMA) and DOSPER (1,3-di-oleoyloxy-2-(6-carboxy-spermyl)-propyl-amide tetraacetate, commercially available from Boehringer-Mannheim). Additional examples of lipids that can be used in the invention can be found in Gao, X. and Huany, L., 1995, *Gene Therapy* 2, 710. Saccharide ligands can be added in the form of saccharide cerebrosides, e.g., lactosylcerebroside or galactocerebroside (Avanti Polar Lipids).

The particular choice of lipid is not critical. Hydrogenated EPC or lysolecithin can be used in place of EPC. DPPC (dipalmitoyl phosphatidylcholine), can be incorporated to improve the efficacy and/or stability of the delivery system.

6.3.3 The Construction of Lipid Nanosphere Carriers

Lipid nanospheres can be constructed by the following process. A methanol or chloroform methanol solution of phospholipids is added to a small test tube and the solvent removed by a nitrogen stream to leave a lipid film. A lipophilic salt of CMV is formed by mixing an aqueous saline solution of CMV with an ethanolic solution of a cationic lipid. Good results can be obtained when the cationic species are in about a 4 fold molar excess relative to the CMV anions (phosphates). The lipophilic CMV salt solution is added to the lipid film, vortexed gently followed by the addition of an amount of neutral lipid equal in weight to the phospholipids. The concentration of CMV can be up to about 3% (w/w) of the total amount of lipid.

After addition of the neutral lipid, the emulsion is sonicated at 4°C for about 1 hour until the formation of a milky suspension with no obvious signs of separation. The suspension is extruded through polycarbonate filters until a final diameter of about 50 nm is achieved. When the target cell is a reticuloendothelial cell the preferred diameter of the lipid nanospheres is about 100-200 nm. The CMV-carrying lipid nanospheres can then be washed and placed into a pharmaceutically acceptable carrier or tissue culture medium. The capacity of lipid nanospheres is about 2.5 mg CMV/ 500 μ l of a nanosphere suspension.

6.3.4 The Construction of Lipid Vesicles

A lipid film is formed by placing a chloroform methanol solution of lipid in a tube and removing the solvent by a nitrogen stream. An aqueous saline solution of CMV is added such that the amount of CMV is between 20% and 50% (w/w) of the amount of lipid, and the amount of aqueous solvent is about 80% (w/w) of the amount of lipid in the final mixture. After gentle vortexing the liposome-containing liquid is forced through successively finer polycarbonate filter membranes until a final diameter of about 50 nm is achieved. The passage through the successively finer polycarbonate filter results in the conversion of polylamellar liposomes into unilamellar liposomes, i.e., vesicles. When the target cell is a reticuloendothelial cell the preferred diameter of the lipid nanospheres is about 100-200 nm. The CMV-carrying lipid nanospheres can then be washed and placed into a pharmaceutically acceptable carrier or tissue culture medium.

The CMV are entrapped in the aqueous core of the vesicles. About 50% of the added CMV is entrapped.

A variation of the basic procedure comprises the formation of an aqueous solution containing a PEI/CMV condensate at a ratio of about 4 PEI imines per CMV phosphate. The condensate can be particularly useful when the liposomes are positively charged, i.e., the lipid vesicle contains a concentration of cations of cationic lipids such as DOTAP, DOTMA or DOSPER, greater than the concentration of anions of anionic lipids such as DOPS. The capacity of lipid vesicles is about 150 μ g CMV per 500 μ l of a lipid vesicle suspension.

In a preferred embodiment the lipid vesicles contain a mixture of the anionic phospholipid, DOPS, and a neutral lipid such as DOPE or DOPC. Other negatively charged phospholipids that can be used to make lipid vesicles include dioleoyl phosphatidic acid (DOPA) and dioleoyl phosphatidyl glycerol (DOPG). In a more preferred embodiment the neutral lipid is DOPC and the ratio of DOPS:DOPC is between 2:1 and 1:2 and is preferably about 1:1. The ratio of negatively charged to neutral lipid should be greater than 1:9 because the presence of less than 10% charged lipid results in instability of the lipid vesicles because of vesicle fusion.

A particular lipid vesicle formulation can be tested by using the formulation to transfect a target cell population with a plasmid of about 5.0 kb in length that expresses

some readily detectable product in the transfected target cell. Lipid vesicles can be used to transfect a cell with the plasmid if the plasmid is condensed with PEI at an imine:phosphate ratio of about 9-4:1. The capacity of the lipid vesicle formulation to transfect a cell with a plasmid is indicative of the formulation's capacity to introduce a CMV into a cell and effect a transmutation.

Certain lipids, particularly the polycationic lipids, can be toxic to certain cell lines and primary cell cultures. The formulation of the lipid vesicles should be adjusted to avoid such toxic lipids.

Ligands for clathrin-coated pit receptors can be introduced into the lipid vesicles by a variety of means. Cerebrosides, such as lactocerebroside or galactocerebroside can be introduced into the lipid mixture and are incorporated into the vesicle to produce a ligand for the asialoglycoprotein receptor.

In an alternative embodiment the lipid vesicle further comprise an integral membrane protein that inserts itself into the lipid bilayer of the vesicle. In a specific embodiment the protein is a fusogenic (F-protein) from the virus alternatively termed Sendai Virus or Hemagglutinating Virus of Japan (HVJ). The preparation and use of F-protein containing lipid vesicles to introduce DNA into liver, myocardial and endothelial cells have been reported. See, e.g., U.S. Patent No. 5,683,866, International Application PCT JP97/ 00612 (published as WO 97/31656). See also, Ramani, K., et al., 1996, FEBS Letters **404**, 164-168; Kaneda, Y., et al., 1989, J. Biol. Chem. **264**, 121126-12129; Kaneda, Y., et al., 1989, Science **243**, 375; Dzau, V.J., et al., Proc. Natl. Acad. Sci. **93**, 11421-11425; Aoki, M., et al., 1997, J.Mol.Cardiol. **29**, 949-959.

6.4 DISEASES AND DISEASE-SPECIFIC CMV

The invention can be used to correct any disease-causing mutation, in which the mutation results in the change of one or more nucleotides or in the insertion or deletion of from one to about 30 nucleotides. In a preferred embodiment, the deletion or insertion is of from one to about six nucleotides. The disease-causing mutation is corrected by administering a CMV containing the sequence of the wild type gene that is homologous to the locus of the mutation. The CMV is constructed so that there are regions of homology with the mutant DNA sequence flanking the heterologous region,

i.e, the region of the CMV that contains the portion of the wild-type sequence that is absent from the mutant. When the mutation consists of an insertion, the heterologous region of the CMV is considered to be the point which is homologous to the site of the insertion. Accordingly, the length of the heterologous region of the CMV is deemed to be the length of the insertion in the mutant sequence. Note that the sequence of the CMV is determined by the location of the mutation, however, the sequence of the mutation is not important. Rather, the sequence of a CMV is always the sequence of the wild type gene or a desired related sequence. In each of the sequences that follow the heterologous region is underlined.

A first embodiment of the invention is a CMV that can be used to correct the mutation that causes the von Willebrand's Disease. A CMV to correct this mutation contains the sequence 5'-CTC GGA GAG C CCC CTC GCA-3' (SEQ ID No. 27), the sequence of a mixed ribo-deoxyribo oligonucleobase having the same sequence of bases or a sequence that differs by the substitution of the thymine by uracil, or the sequence of the complement thereof. The tissue in which the von Willebrand's factor gene needs to be corrected is the vascular endothelium.

A further embodiment of the invention is a CMV that can be used to correct the mutation that causes Hemophilia B, which is an A→C substitution at nt 1234 of the human coagulation Factor IX gene. CMV to correct this mutation contains the sequence 5'-CAA GGA GAT AGT GGG GGA C-3' (SEQ ID No. 28), the sequence of a mixed ribo-deoxyribo oligonucleobase having the same sequence of bases or a sequence that differs by the substitution of the thymine by uracil, or the sequence of the complement thereof. The invention can be used to correct other mutations in the human coagulation Factor IX gene, the sequence of which is given in Kurachi, K., et al., 1982, Proc. Natl. Acad. Sci. U.S.A. 79, 6461-6464, which is hereby incorporated by reference. The tissue in which the factor IX gene needs to be corrected is the hepatocellular liver.

A further embodiment of the invention is a CMV that can be used to correct the Z-mutation that causes α 1-antitrypsin deficiency. The Z mutation is a G→A substitution located at nt 1145 of the human α 1-antitrypsin gene. CMV to correct this mutation contains the sequence 5'-ACC ATC GAC GAG AAA GGG A-3' (SEQ ID No. 29), the sequence of a mixed ribo-deoxyribo oligonucleobase having the same sequence of bases

or a sequence that differs by the substitution of the thymine by uracil, or the sequence of the complement thereof. The invention can be used to correct other mutations in the α 1-antitrypsin gene, the sequence of which is given in Long, G.L., et al., 1984, *Biochemistry* 23, 4828-4837, which is hereby incorporated by reference. The tissue in which the α 1-antitrypsin gene needs to be corrected is the hepatocellular liver.

A further embodiment of the invention is a CMV that can be used to correct a mutation in the low density lipoprotein receptor (LDLR) that causes familial hypercholesterolemia (FH). There is no single mutation that causes the majority of FH cases. Surveys of the more than 105 point mutations or insertions or deletions of 25 nt or fewer that cause FH can be found in Hobbs, H.H., et al., 1992, *Hum. Mutat.* 1, 445-466 and Leren, T.P., et al., *Hum. Genet.* 95, 671-676, which are hereby incorporated by reference in their entirety. The complete sequence of the human LDLR cDNA is published in Yamamoto, T., et al., 1984, *CELL* 39, 27-38. The tissue in which the LDLR can be corrected to obtain amelioration of FH is the hepatocellular liver.

A further embodiment of the invention is a CMV that can be used to correct a mutation in the glucocerebrosidase gene that causes Gaucher Disease. The structure of CMV that can be used to correct a Gaucher Disease mutation can be found in commonly licensed U.S. application Serial No. 08/640,517. The tissue in which the glucocerebrosidase mutation can be corrected to obtain amelioration of Gaucher Disease is the reticuloendothelial (Kupffer Cell) liver.

A further embodiment of the invention is a CMV that can be used to correct a mutation in the glucose-6-phosphatase (G-6-P) gene that causes type 1 Glycogen Storage Disease (GSD). The complete sequence of the human G-6-P is given in Lei, K.-J., et al., *SCIENCE* 262, 580, which is hereby incorporated by reference. The two most common mutations that cause type 1 GSD are C→T at nt 326, C→T at nt 1118, and an insertion of TA at nt 459, as described in Lei, K.-J., et al., *J. Clin. Investigation* 95, 234-240, which is hereby incorporated by reference. CMV to correct the two most common mutations contain the sequence 5'-TTT GGA CAG CGT CCA TAC T-3' (SEQ ID No. 30), or 5'-TGC CTC GCC CAG GTC CTG G-3' (SEQ ID No. 31), the sequence of a mixed ribo-deoxyribo oligonucleobase having the same sequence of bases or a sequence that differs by the substitution of the thymine by uracil, or the sequence of the complement thereof.

A further embodiment of the invention is a CMV that can be used to correct a mutation in the Ornithine Transcarbamylase (OTC) gene an X-linked gene that catalyzes the condensation of ornithine and carbamyl phosphate to yield citruline and phosphate. The complete sequence of the human OTC cDNA is given in Horwich, A.L., et al., 1984, *Science* **224**, 1068, which is hereby incorporated by reference. The structure of OTC gene and a review of the structure of identified mutants is reviewed in Tuchman, M., 1992, *Human Mutation* **2**, 174.

A further embodiment of the invention is a CMV that can be used to correct a mutation in the human UDP-glucuronosyltransferase gene that causes Crigler-Najjar syndrome. The sequence of the human UDP-glucuronosyltransferase gene is given in Bosma, P.J., et al., 1992, *Hepatology*, **15**, 941-7, which is hereby incorporated by reference. The tissue in which the UDP-glucuronosyltransferase gene can be corrected to obtain amelioration of Crigler-Najjar syndrome is the hepatocellular liver.

A further embodiment of the invention is a CMV that can be used to correct a mutation in a galactose-1-phosphate uridyltransferase gene that cause galactosemia. The sequence of the human galactose-1-phosphate uridyltransferase gene is described in Flack, J.E., et al., 1990 *Mol. Biol. Med.* **7**, 365, and the molecular biology and population genetics of galactosemia are described in Reichert, J.K.V., et al., 1991, *Proc. Natl. Acad. Sci.* **88**, 2633-37 and Reichert, J.K.V., et al., 1991, *Am. J. Hum. Gen.* **49**, 860, which are hereby incorporated by reference. The most common mutation that causes galactosemia is Q→R at amino acid 188. The CMV to correct this mutation contains the sequence 5'-CC CAC TGC CAG GTA TGG GC-3' (SEQ ID No. 32), the sequence of a mixed ribo-deoxyribo oligonucleobase having the same sequence of bases or a sequence that differs by the substitution of the thymine by uracil, or the sequence of the complement thereof.

A further embodiment of the invention is a CMV that can be used to correct a mutation in the phenylalanine hydroxylase (PAH, phenylalanine 4-monooxygenase, EC 1.14.16.1) that cause phenylketonuria (PKU) or hyperphenylalaninemia. The molecular and population genetics of phenylketonuria are described in Woo, SLC, 1989, *Biochemistry* **28**, 1-7, the sequence of human PAH is described in Kowk, S.C.M., et al., 1985, *Biochemistry* **28**, 556-561, which are hereby incorporated by reference. Further

examples of PKU-causing mutations can be found in Sworniczak, B., et al., 1992, Hum. Mutat. 1, 138-146.

6.5 THE USE OF THE FORMULATIONS *IN VIVO*

The CMV of the invention can be parenterally administered directly to the target organ at a dose of between 50 and 250 $\mu\text{g/gm}$. When the target organ is the liver muscle or kidney, the CMV/macromolecular carrier complex can be injected directly into the organ. When the target organ is the liver, intravenous injection into the hepatic or portal veins of a liver, having temporarily obstructed circulation can be used. Alternatively the CMV/macromolecular complex can further comprise a hepatic targeting ligand, such as a lactosyl or galactosyl saccharide, which allows for administration of the CMV/macromolecular complex intravenously into the general circulation.

When the target organ is the lung or a tissue thereof, e.g., the bronchiolar epithelium CMV/macromolecular complex can be administered by aerosol. Small particle aerosol delivery of liposomal/DNA complexes is described in Schwarz L.A., et al., 1996, Human Gene Therapy 7, 731-741.

When the target organ is the vascular endothelium, as for example in von Willebrand's Disease, the CMV/macromolecular complex can be delivered directly into the systemic circulation. Other organs can be targeted by use of liposomes that are provided with ligands that enable the liposome to be extravasated through the endothelial cells of the circulatory system.

For enzymatic defects, therapeutic effects can be obtained by correcting the genes of about 1% of the cells of the affected tissue. In a tissue in which the parenchymal cells have an extended life, such as the liver, treatments with CMV can be repeatedly performed to obtain an increased therapeutic effect.

7. EXAMPLES

7.1 CMV/MACROMOLECULAR CARRIER COMPLEXES

7.1.1 Lipid Nanospheres

Materials

Egg phosphatidylcholine (EPC), DOTAP and galactocerebroside (Gc) (Avanti Polar Lipids); soybean oil (Sigma Chemical Co.); dioctadecyldiamidoglycyl spermine (DOGS®) (Promega).

Methods

EPC, DOTAP and Gc were previously dissolved at defined concentrations in chloroform or anhydrous methanol and stored in small glass vials in desiccated containers at -20°C until use. EPC (40-45 mg), DOTAP (200 µg) and Gc (43 µg) solutions were aliquoted into a small 10 x 75 mm borosilicate tube and solvents removed under a stream of nitrogen. CMV were diluted in 0.15 M NaCl (~80-125 µg/250-300 µl); DOGS (as a 10 mg/ml solution in ethanol) was diluted into 250-300 µl 0.15 M NaCl at 3-5 times the weight of added CMV. The two solutions were mildly vortexed to mix contents and then CMV solution was added slowly to the DOGS solution. The contents were mixed by gentle tapping and inverting the tube a few times. The DOGS-complex solution was added to the dried lipids followed by soybean oil (40-45 mg), the mixture was vortexed on high for a few seconds and bath sonicated in a FS-15 (Fisher Scientific) bath sonicator for ~1 hr in a 4°C temperature controlled room. Occasionally, the tube was removed from the bath and vortexed. When a uniform looking, milky suspension was formed (with no obvious separation of oil droplets), it was extruded through a series of polycarbonate membranes down to a pore size of 50 nm. Preparations were stored at 4°C until use and vortexed before use.

7.1.2 Negatively Charged Targeted Lipid Vesicles

Materials

Dioleoyl phosphatidylcholine (DOPC), dioleoyl phosphatidylserine (DOPS), galactocerebroside (Gc) or lactosylcerebroside, (Avanti Polar Lipids).

Methods

DOPS, DOPC and Gc at a molar ratio of 1:1:0.16 (500 μ g total lipid) were dissolved in chloroform:methanol (1:1 v/v) and then dried under a stream of nitrogen to obtain a uniform lipid film. The CMV were diluted in 500 μ l of 0.15 M NaCl (approximately 100-250 μ g/500 μ l). The solution was added to the lipid film at room temperature. Lipids were dispersed entirely by alternate mild vortexing and warming (in a water bath at 37-42°C). After a uniform milky suspension was formed, it was extruded through a series of polycarbonate membranes (pore sizes 0.8, 0.4, 0.2, 0.1 and 0.05 μ m) using a Liposofast® mini-extruder. Extrusion was done 5 to 7 times through each pore size. After preparation, lipid vesicles were stored at 4°C until use. Under these conditions the lipid vesicles were stable for at least one month. The final product can be lyophilized.

7.1.3 Neutral Targeted Lipid Vesicles

Materials

Dioleoyl phosphatidylcholine (DOPC), dioleoyl phosphatidylethanolamine (DOPE), galactocerebroside (Gc) or lactosylcerebroside, (Avanti Polar Lipids).

Methods

DOPC, DOPE and Gc (1:1:0.16 molar ratio) or DOPC:Gc (1:0.08) were dissolved in chloroform:methanol (1:1 v/v) and then dried under a stream of nitrogen to obtain a uniform lipid film. The oligonucleotides (or chimeric molecules) were diluted in 500 μ l of 0.15 M NaCl (approximately 100-250 μ g/500 μ l). The solution was added to the lipid film at room temperature. Lipids were dispersed entirely by alternate mild vortexing and warming (in a water bath at 37-42°C). After a uniform milky suspension was formed, it was extruded through a series of polycarbonate membranes (pore sizes 0.8, 0.4, 0.2, 0.1 and 0.05 μ m) using a Liposofast® mini-extruder. Extrusion was done 5 to 7 times through each pore size. After preparation, lipid vesicles were stored at 4°C until use. The size of the lipid vesicles of the preparation was stable for about 5 days.

7.1.4 Positively Charged Targeted Lipid Vesicles

Materials

Dioleoyl phosphatidylcholine (DOPC), dioleoyl trimethylammonium propane (DOTAP), galactocerebroside (Gc) or lactosylcerebroside, (Avanti Polar Lipids). Polyethylenimine (PEI) (M.W. 800 Kd), Fluka Chemicals.

Methods

DOPC, DOTAP and Gc (6:1:0.56 molar ratio) (500 μg total lipid) were dissolved in chloroform:methanol (1:1 v/v) and then dried under a stream of nitrogen to obtain a uniform lipid film. PEI was diluted to a concentration of 45 mg/100 ml using water. pH of the solution was adjusted to ~ 7.6 using HCl. This PEI stock solution was prepared fresh each time and was equivalent to approximately 50 nmol amine/ μl . CMV were diluted into 0.15 M NaCl at a concentration of $\sim 125 \mu\text{g}$ in 250 μl . PEI was further diluted into 250 μl 0.15 M NaCl so that approximately 4 moles of PEI amine were present per mole of oligonucleotide/chimeric phosphate. PEI solution was added dropwise to the CMV solution (both at room temperature) and vortexed for 5-10 minutes. The PEI-complex solution was then added to the lipid film and the lipids dispersed as described above. After a uniform milky suspension was formed, it was extruded through a series of polycarbonate membranes (pore sizes 0.8, 0.4, 0.2, 0.1 and 0.05 μm) using a Liposofast® mini-extruder. Extrusion was done 5 to 7 times through each pore size. After preparation lipid vesicles were stored at 4°C until use. Under these conditions the lipid vesicles were stable for at least one month. For longer and improved stability the final product can be lyophilized.

7.1.5 Lactosylated-PEI/PEI Complexes

PEI (25 kDa) was purchased from Aldrich Chemical (Milwaukee, WI). PEI (800 kDa) was purchased from Fluka chemicals (Ronkonkoma, NY, USA). Lactosylation of the PEI was carried out by modification of a previously described method for the conjugation of oligosaccharides to proteins. Briefly, 3 to 5 ml of PEI (0.1 to 1.2 M_{monomer}) in ammonium acetate (0.2 M) or Tris buffer (0.2 M) (pH 7.6) solution was incubated with 7

to 8 mg of sodium cyanoborohydride (Sigma Chemical Co., St. Louis, MO) and approximately 30 mg of lactose monohydrate (Sigma Chemical Co., St. Louis, MO). Reaction was carried out in polypropylene tubes, tightly capped in a 37°C shaking water bath. After 10 days the reaction mixture was dialyzed against distilled water (500 ml) for 48 h with 1 to 2 changes of water. The purified complex was sterile filtered through 0.2 μ m filter and stored at 4°C. The amount of sugar (as galactose) associated with PEI was determined by the phenol-sulphuric acid method.

The number of moles of free amine (primary + secondary) in the lactosylated PEI was determined as follows: a standard curve was set up using a 0.02M stock solution of PEI; several aliquots of the stock were diluted to 1ml using deionized water in glass tubes, then 50 μ l of Ninhydrin reagent (Sigma Chemical Co., St. Louis, Mo) was added to each tube and vortexed vigorously for 10 sec. Color development was allowed to proceed at room temperature for 10 to 12 min. and then O.D. was read (within 4 minutes) at 485 nm on a Beckman DU-64 spectrophotometer. 20 to 50 μ l aliquots of the L-PEI samples were treated as above and the number of moles of free amine was determined from the standard curve. Lactosylated-PEI (L-PEI) complexes were prepared as follows: an equivalent of 3 mmol of amine as L-PEI and 3 mmol of amine as PEI, per mmol of RNA/DNA phosphate, were mixed together and diluted in 0.15M NaCl as required; the mixture was added dropwise to a solution of the chimeric and vortexed for 5 min.

To verify complete association of the chimeric oligonucleotides with PEI or L-PEI, gel analysis (4% LMP agarose) of the uncomplexed and complexed chimerics was performed. To determine the degree of protection against nuclease degradation provided by complexation of the chimerics, samples were treated with RNase and DNase. After a chloroform phenol extraction, the complexes were dissociated using heparin (50 units/ μ g nucleic acid) and the products analyzed on a 4% LMP agarose gel.

7.2 DEMONSTRATION OF PEI/CMV MEDIATED ALTERATION OF RAT AND HUMAN FACTOR IX

Materials. Fetal bovine serum was obtained from Atlanta Biologicals, Inc. (Atlanta, GA). The terminal transferase, fluorescein-12-dUTP, Expand™ high fidelity PCR system, dNTPs and high pure PCR template preparation kit were obtained from

Boehringer Mannheim Corp. (Indianapolis, IN). Reflection™ NEF-496 autoradiography film and Reflection™ NEF-491 intensifying screens were from DuPont NEN® Research Products (Boston, MA). Polyethylenimine (PEI) 800 kDa was obtained from Fluka Chemical Corp. (Ronkonkoma, NY). The [γ -³²P]ATP was obtained from ICN Biochemicals, Inc. (Costa Mesa, CA). pCR™2.1 was obtained from Invitrogen (San Diego, CA). OPTIMEM™, Dulbecco's modified Eagle's medium, William's E medium and oligonucleotides 365-A and 365-C were from Life Technologies, Inc. (Gaithersburg, MD). Spin filters of 30,000 mol wt cutoff were purchased from Millipore Corp. (Bedford, MA). Dil and SlowFade™ antifade mounting medium were obtained from Molecular Probes, Inc. (Eugene, OR). T4 polynucleotide kinase was purchased from New England Biolabs, Inc. (Beverly, MA). MSI MagnaGraph membrane was purchased from Micron Separations, Inc. (Westboro, MA). The primers used for PCR amplification were obtained from Oligos Etc., Inc. (Wilsonville, OR). Tetramethylammonium chloride was purchased from Sigma Chemical Company (St. Louis, MO). All other chemicals were molecular biology or reagent grade and purchased from Aldrich Chemical Company (Milwaukee, WI), Curtin Matheson Scientific, Inc. (Eden Prairie, MN), and Fisher Scientific (Itasca, IL).

Oligonucleotide synthesis. Chimeric RNA/DNA oligonucleotides HIXF, RIXF and RIXR were synthesized. The CMV were prepared with DNA and 2'-O-methyl RNA phosphoramidite nucleoside monomers on an ABI 394 synthesizer. The DNA phosphoramidite exocyclic amine groups were protected with benzoyl (adenosine and cytidine) and isobutyryl (guanosine). The protective groups on the 2'-O-methyl RNA phosphoramidites were phenoxyacetyl for adenosine, isobutyryl for cytidine, and dimethylformamide for guanosine. The base protecting groups were removed following synthesis by heating in ethanol/concentrated ammonium hydroxide for 20 h at 55°C. The crude oligonucleotides were electrophoresed on 15% polyacrylamide gels containing 7 M urea, and the DNA visualized using UV shadowing. The chimeric molecules were eluted from the gel slices, concentrated by precipitation and desalted using G-25 spin columns. Greater than 95% of the purified oligonucleotides were full length.

The sequence of the wild type and "mutant" rat Factor IX are

(SEQ ID No. 33)

365

wt AAA GAT TCA TGT GAA GGA GAT AGT GGG GGA CCC CAT GTT
Lys Asp Ser Cys Glu Gly Asp Ser Gly Gly Pro His Val

(SEQ ID No. 34)

(SEQ ID No. 35)

mt AAA GAT TCA TGT GAA GGA GAT CGT GGG GGA CCC CAT GTT
Arg

The structure of the RIXR, RIXF and HIXR CMV is as follows:

Chimeric Oligonucleotides

RIXR

(SEQ ID No. 36)

TGCGCG-ccccagggggTGCTAgaggaaguguT
T T
T T
TCGCGC GGGGTCCCCCACGATCTCCTTCACAT
3' 5'

RIXR_C

(SEQ ID No. 37)

TGCGCG-acacuuccucTAGCAccccccuggggT
T T
T T
TCGCGC TGTGAAGGAGATCGTTGGGGGACCCCT
3' 5'

RIXF

(SEQ ID No. 38)

TGCGCG-acacuuccucTAGCAccccccuggggT
T T
T T
TCGCGC TGTGAAGGAGATCGTTGGGGGACCCCT
3' 5'

HIXF (SEQ ID No. 39)

TGCGCG-acaguuccucTAGCAccccccuggggT

T T

T T

TCGCGC TGTCAAGGAGATTCGTGGGGGACCCCT

3' 5'

Uppercase letters are deoxyribonucleotides, lower case letters are 2'OMe-ribonucleotides. The nucleotide of the heterologous region is underlined.

Cell Culture, transfections and hepatocyte isolation. HuH-7 cells were maintained in Dulbecco's modified Eagle's medium containing 10% (vol/vol) heat inactivated fetal bovine serum in a humidified CO₂ atmosphere at 37°C. Twenty four hours prior to transfection 1 x 10⁵ cells were plated per 35 mm culture dish. At the time of transfection, the cells were rinsed twice with OPTIMEM™ media and transfections were performed in 1 ml of the same media. Eighteen hours after transfection, 2 ml of Dulbecco's modified Eagle's medium containing 20% (vol/vol) heat inactivated fetal bovine serum was added to each 35 mm dish and the cells maintained for an additional 30 h prior to harvesting for DNA isolation. A PEI (800 kDa) 10 mM stock solution, pH 7.0, was prepared. Briefly, the chimeric oligonucleotides were transfected with 10 mM PEI at 9 equivalents of PEI nitrogen per chimeric phosphate in 100 µl of 0.15 M NaCl at final concentrations of either 150 nM (4 µg), 300 nM (8 µg) and 450 nM (12 µg). After 18 h, an additional 2 ml of medium was added and reduced the chimeric concentrations to 50 nM, 100 nM, and 150 nM, respectively, for the remaining 30 h of culture. HuH-7 vehicle control transfections utilized the same amount of PEI as was used in the HuIXF transfections, but substituted an equal volume of 10 mM Tris-HCl pH 7.6 for the oligonucleotides.

Primary rat hepatocytes were isolated from 250 g male Sprague-Dawley rats (Harlan Sprague-Dawley, Inc., Indianapolis, IN) by a two step collagenase perfusion as previously described (Fan et al., Oncogene 12:1909-1919, 1996, which is hereby incorporated by reference) and plated on Primaria™ plates at a density of 4 x 10⁵ cells per 35 mm dish. The cultures were maintained in William's E medium supplemented with 10% heat inactivated FBS, 26 mM sodium bicarbonate, 23 mM HEPES, 0.01 U/ml

insulin, 2 mM L-glutamine, 10 nM dexamethasone, 5.5 mM glucose, 100 U/ml penicillin and 100 U/ml streptomycin. Twenty four hours after plating, the hepatocytes were washed twice with the same medium and 1 ml of fresh medium added and the cells transfected using PEI/chimeric oligonucleotide complexes at the identical concentrations as for the HuH-7 cells. After 18 h, an additional 2 ml of the medium was added and the cells harvested 6 or 30 h later.

Direct injection of chimeric oligonucleotides into liver. Male Sprague-Dawley rats (~175 g) were maintained on a standard 12 h light-dark cycle and fed *ad libitum* standard laboratory chow. The rats were anesthetized, a midline incision made the liver exposed. A clamp was placed on the hepatic and portal veins as they enter the caudate lobe, and 75 μ g of the 1:9 chimeric/PEI complex was injected in a final volume of 250 - 300 μ l directly into the caudate lobe. The lobe remained ligated for 15 min and then blood flow was restored by removing the clamp. After suturing the incision the animals were allowed to recover from the anesthesia and given food and water *ad libitum*. Vehicle controls were done substituting an equal volume of Tris-HCl pH 7.6 for the chimeric oligonucleotides. Twenty-four and 48 h post-injection the animals were sacrificed, the caudate lobe removed and the tissue around the injection site dissected for DNA isolation. DNA was isolated and the terminal exon of the rat factor IX gene was amplified by PCR.

Nuclear uptake of the chimeric molecules. Chimeric duplexes were 3' end-labeled using terminal transferase and fluorescein-12-dUTP according to the manufacturer's recommendation, and were then mixed with unlabeled oligonucleotides at a 2:3 ratio. Transfections were performed as described above and after 24 h the cells were fixed in phosphate buffered saline, pH 7.4, containing 4% paraformaldehyde (wt/vol) for 10 min at room temperature. Following fixation, the cells were counterstained using a 5 μ M solution of Dil in 0.32 M sucrose for 10 min according to the manufacturer's recommendation. After washing with 0.32 M sucrose and then phosphate buffered saline, pH 7.4, the cells were coverslipped using SlowFade™ antifade mounting medium in phosphate buffered saline and examined using a MRC1000 confocal microscope (BioRad, Inc., Hercules, CA). The caudate lobes of liver *in situ* were injected with fluorescently-labeled chimerics as described above and harvested 24 h post-

injection. The lobes were bisected longitudinally, embedded using OCT and frozen. Cryosections were cut ~10 μ m thick, fixed for 10 min at room temperature using phosphate buffered saline, pH 7.4, containing 4% paraformaldehyde (wt/vol). Following fixation, the cells were counterstained using a 5 μ M solution of Dil in 0.32 M sucrose for 10 min according to the manufacturer's recommendation. After washing with 0.32 M sucrose and then phosphate buffered saline, pH 7.4, the sections were coverslipped using SlowFade™ antifade mounting medium and examined using a MRC1000 confocal microscope (BioRad, Inc.). The collection series for the fixed cells and sectioned tissue were made at 1 μ m steps to establish the presence of the chimeric in the nucleus.

DNA isolation and cloning. The cells were harvested by scrapping 24 and 48 h after transfection. Genomic DNA larger than 100-150 base pairs was isolated using the high pure PCR template preparation kit according to the manufacturer's recommendation. PCR amplification of a 317-nt fragment of the eighth exon in the human liver factor IX gene was performed with 500 ng of the isolated DNA. The primers used were 5'-CATTGCTGACAAGGAATACACGAAC-3' (SEQ ID No. 40) and 5'-ATTTGCCTTTCATTGCACACTCTTC-3' (SEQ ID No. 41) corresponding to nucleotides 1008-1032 and 1300-1324, respectively, of the human factor IX cDNA. Primers were annealed at 58°C for 20 sec, elongation was for 45 sec at 72°C and denaturation proceeded for 45 sec at 94°C. The sample was amplified for 30 cycles using Expand Hi-fidelity™ polymerase. PCR amplification of a 374-nt fragment of the rat factor IX gene was performed with 500 ng of the isolated DNA from either the primary hepatocytes or liver caudate lobe. The primers used were 5'-ATTGCCTTGCTGGAAGTGGATAAC-3' (SEQ ID No. 42) and 5'-TTGCCTTTCATTGCACATTCTTCAC-3' (SEQ ID No. 43) corresponding to nucleotides 433-457 and 782-806, respectively, of the rat factor IX cDNA. Primers were annealed at 59°C for 20 sec, elongation was for 45 sec at 72°C and denaturation proceeded for 45 sec at 94°C. The sample was amplified for 30 cycles using Expand Hi-fidelity™ polymerase. The PCR amplification products from both the human and rat factor IX genes were subcloned into the TA cloning vector pCR™2.1 according to the manufacturer's recommendations, and the ligated material used to transform frozen competent *Escherichia coli*.

Colony hybridization and sequencing. Eighteen to 20 h after plating, the colonies were lifted onto MSI MagnaGraph nylon filters, replicated and processed for hybridization according to the manufacturer's recommendation. The filters were hybridized for 24 h with 17 mer oligonucleotide probes 365-A (5'-AAGGAGATAGTGGGGGA-3') (SEQ ID No. 44) or 365-C (5'-AAGGAGATCGTGGGGGA-3') (SEQ ID No. 45), where the underlined nucleotide is the target of the mutagenesis. The probes were ³²P-end-labeled using [γ-³²]ATP (>7,000 Ci/mmol) and T4 polynucleotide kinase according to the manufacturer's recommendations. Hybridizations were performed at 37°C in 2X sodium chloride sodium citrate containing 1% SDS, 5X Denhardt's and 200 µg/ml denatured sonicated fish sperm DNA. After hybridization, the filters were rinsed in 1X sodium chloride phosphate EDTA, 0.5% SDS and then washed at 54°C for 1 h in 50 mM Tris-HCl, pH 8.0 containing 3 M tetramethylammonium chloride, 2 mM EDTA, pH 8.0, 0.1% SDS. Autoradiography was performed with NEN® Reflection film at -70°C using an intensifying screen. Plasmid DNA was prepared from colonies identified as hybridizing with 365-A or 365-C using Qiagen miniprep kit (Chatsworth, CA) and subjected to automatic sequencing using the mp13 reverse primer on an ABI 370A sequencer (Perkin-Elmer, Corp., Foster City, CA).

Results *In Vivo*

Chimeric oligonucleotides were fluorescein-labeled and used to determine whether direct injection into the caudate lobe of the liver was feasible. The results indicated that the hepatocytes adjacent to the injection site within the caudate lobe showed uptake of the fluorescently-labeled chimeric molecules similar to that observed in isolated primary hepatocytes and HuH-7 cells. Although some punctate material was present in the cytoplasm, the labeled material was detected primarily in the nucleus. In fact, only nuclear labeling was observed in hepatocytes farthest from the injection site. The unlabeled PEI/RIXF chimeric complexes and vehicle controls were injected directly into the caudate lobe using the same protocol and the animals sacrificed 24 and 48 h post-injection. Liver DNA was isolated as described in Methods, subjected to PCR amplification of a 374 nt sequence spanning the targeted nt exchange site. Following

subcloning and transformation of *Escherichia coli* with the PCR amplified material, duplicate filter lifts of the transformed colonies were performed. The filters were hybridized with ³²-labeled 17-mer oligonucleotides specific for either 365-A (wild-type) or 365-C (factor IX mutation) and processed post-hybridization as described in Methods. Rats which received direct hepatic injection of the RIXF chimeric molecules exhibited a A→C conversion frequency of ~1% at both 24 and 48 h. In contrast, the vehicle controls showed no hybridization with the 365-C probe. Colonies that hybridized with the 365-C probe from the RIXF treated animals were cultured, the plasmid DNA isolated and subjected to sequencing to confirm the A→C conversion. The ends of the amplified 374-nt fragment correspond exactly with the primers and the only nucleotide change observed was an A→C at the targeted exchange site.

7.3 DEMONSTRATION OF LACTOSYLATED-PEI/CMV MEDIATED ALTERATION OF RAT FACTOR IX

7.3.1 Results

CMV complexed to a mixture of lactosylated-PEI and PEI was prepared using the RIXR oligonucleotide as described in Section 6.1.5 above. A CMV directed to the complementary strand of the same region of the factor IX was also constructed (RIXR_C).

Conversion of the targeted nucleotide at Ser³⁶⁵ by the chimeric oligonucleotides

The nuclear localization of the fluorescently-labeled chimeric molecules indicated efficient transfection in the isolated rat hepatocytes. The cultured hepatocytes were then transfected with the unlabeled chimeric molecules factor RIXR_C and RIXR at comparable concentrations using 800 kDa PEI as the carrier. Additionally, vehicle control transfections were performed simultaneously. Forty eight hours after transfection, the cells were harvested and the DNA isolated and processed for hybridization as described in Section 6.1.5. The A→C targeted nucleotide conversion at Ser³⁶⁵ was determined by hybridization of duplicate colony lifts of the PCR-amplified and cloned 374-nt stretch of exon 8 of the factor IX gene (Sarkar, B., Koeberl, D. D. & Somer, S. S., "Direct Sequencing of the activation peptide and the catalytic domain of the factor IX gene in six species," *Genomics*, 6, 133-143, 1990.) The 17 mer oligonucleotide probes used to

distinguish between the wild-type 365-A (5'-AAGGAGATAGTGGGGGA-3') (SEQ ID No. 46) or converted 365-C (5'-AAGGAGATCGTGGGGGA-3') (SEQ ID No. 47) corresponded to nucleotides 710 through 726 of the cDNA sequence.

The overall frequency of conversion of the targeted nucleotide was calculated by dividing the number of clones hybridizing with the 365-C oligonucleotide by the total number of clones hybridizing with both oligonucleotide probes. The results are summarized in Table III for RIXR_C. A→C conversion at Ser³⁶⁵ was observed only in primary hepatocytes transfected with the RIXR or RIXR_C. Similar conversion frequencies were observed in hepatocytes transfected with RIXR or RIXR_C. Neither vehicle transfected cells nor those transfected with other chimeric oligonucleotides yielded any clones hybridizing with the 365-C oligonucleotide probe (unpublished observations). Additionally, no hybridization of the 365-C oligonucleotide probe was observed to clones derived from DNA isolated from untreated hepatocytes and PCR-amplified in the presence of 0.5 to 1.5 µg of the oligonucleotides. The A→C conversion rate in the isolated hepatocytes was also dose dependent using lactosylated PEI derivatives as described in Section 6.1.5 and was as high as 19%. RT-PCR and hybridization analysis of RNA isolated from cultured cells transfected in parallel with lactosylated PEIs demonstrated A→C conversion frequencies ranging from 11.9 to 22.3%.

Site-directed nucleotide exchange by chimeric oligonucleotides in intact liver

The fluorescein-labeled oligonucleotides were also used to determine cellular uptake of the chimeric molecules after direct injection into the caudate lobe of the liver. The results indicated that hepatocytes adjacent to the injection site in the caudate lobe showed uptake of the fluorescently-labeled chimerics similar to that observed in the isolated rat hepatocytes. Although some punctate material was present in the cytoplasm of the hepatocytes, the labeled material was primarily present in the nucleus. In fact, only nuclear labeling was observed in those areas farthest from the injection site. The unlabeled RIXR chimeric oligonucleotides and vehicle controls were then administered *in vivo* by tail vein injection of the 25 kDa PEI and liver tissue harvested 5 days post-injection. Liver DNA was isolated and subjected to PCR amplification of a 374-nt sequence spanning the targeted nucleotide exchange site, using the same primers as those

used with the primary hepatocytes. Following subcloning and transformation of *E. coli* with the PCR-amplified material, duplicate filter lifts of the transformed colonies were done. The filters were hybridized with the same ^{32}P -labeled 17-mer oligonucleotides specific for either 365-A (wild-type) or 365-C (mutant) and processed post-hybridization. Rats treated with 100 μg of the RIXR chimeric oligonucleotides exhibited an A \rightarrow C conversion frequency ranging from 13.9% to 18.9%, while those that received a total of 350 μg in two injections showed 40% conversion. In contrast, the vehicle controls showed no hybridization with the 365-C probe. RT-PCR hybridization of isolated RNA indicated A \rightarrow C conversion frequencies of 26.4% to 28.4% in the high dose livers. The APTT for vehicle-treated rats ranged from 89.7% to 181.9% of control values ($131.84\% \pm 32.89\%$), while the APTT for the oligonucleotide-treated animals ranged from 48.9% to 61.7% ($53.8\% \pm 4.8\%$).

The APTT times for a 1/10 dilution of rat test plasma in Hepes buffer (50 mM Hepes/100 mM NaCl/0.02% NaN_3 , pH 7.4) were determined for both normal ($n = 9$) and the double injected animals ($n = 3$). The factor IX activity of duplicate samples was determined from a log-log standard curve that was constructed from the APTT results for dilution (1:10 to 1:80) of pooled plasma from 12 normal male rats, 6-8 weeks old. The APTT results for the normal rats ranged from 89.7% to 181.9% of the control values (mean = $131.84\% \pm 32.89\%$), while the APTT results for the double injected animals ranged from 49.0% to 61.7% (mean $53.8\% \pm 5.8\%$). The APTT clotting time in seconds for the normal rats ranged from 60.9 seconds to 81.6 seconds (mean = 71.3 ± 7.3 seconds) while the APTT times ranged from 92.3 seconds to 98.6 seconds (mean = 96.3 ± 2.9 seconds) for the double-infected rats.

Sequence analysis of the mutated factor IX gene in isolated hepatocytes and intact liver

Direct sequencing of the wild-type and mutated genes was performed to confirm the results from the filter hybridizations in both the *in vitro* and *in vivo* studies. At least 10 independent clones hybridizing to either 365-A or 365-C from the intact liver or isolated hepatocytes were analyzed. The results of the sequencing indicated that colonies hybridizing to 365-A (Fig. 6, top panel) exhibited the wild-type IX sequence, i.e. an A at Ser³⁶⁵ of the reported cDNA sequence. In contrast, those colonies derived from the factor

R1XR_C transfected primary hepatocytes hybridizing to the 365-C oligonucleotide probe converted to a C at Ser³⁶⁵. The same A→C conversion at Ser³⁶⁵ was observed in the clones derived from the transfected rat liver that hybridized with the 17 mer 365-C oligonucleotide probe. The entire 374-nt PCR amplified region of the factor IX gene was sequenced for all the clones and no alteration other than the indicated changes at Ser³⁶⁵ was detected. Finally, the start and end points of the 374-nt PCR amplified genomic DNA derived from both the primary hepatocytes and the intact liver corresponded exactly to those of the primers used for the amplification process, indicating that the cloned and sequenced DNA was derived from genomic DNA rather than nondegraded chimeric oligonucleotides.

Table III Percent A→C conversion at Ser³⁶⁵ of rat factor IX genomic DNA by colony lift hybridizations

PEI Deliver System		365-C clones	Total clones	A→C (%)
PEI 800 kDa ¹	<u>Concentration</u>			
<i>In vitro</i>	150 nM	24	572	4.2
	300	31	367	8.5
	450	63	502	12.5
Lac-PEI 800 kDa				
<i>In vitro</i>	90	18	337	5.3
	180	34	300	11.3
	270	47	253	18.6
Lac-PEI 25 kDa				
<i>In vitro</i>	90	28	527	5.3
	180	53	417	12.7
	270	60	305	19.7
Lac-PEI 25 kDa ²	<u>Dose</u>			
<i>In vivo</i> x1	100 µg	24	166	14.5
		71	386	18.4
		50	360	13.9
Lac-PEI 25 kDa				
<i>In vivo</i> x2	350 µg	237	601	39.4
		228	563	40.5
		271	678	40.0

¹The data shown for the primary hepatocyte transfections represents a mean of two experiments.

²The *in vivo* chimeric/PEI complexes were administered in a volume of 300 µl of 5% dextrose by tail vein injection. The results of three animals at each dose are shown individually.

7.3.2 Materials and Methods

In vivo delivery of the chimeric oligonucleotides. Male Sprague-Dawley rats (Harlan Sprague-Dawley, Inc.) (~50 g) were maintained on a standard 12 h light-dark cycle and fed *ad libitum* standard laboratory chow. Vehicle controls and lactosylated 25 kDa PEI at a ratio of 6 equivalents of PEI nitrogen per chimeric phosphate were administered in 300 μ l of 5% dextrose (Abdallah, B. et al., "A powerful nonviral vector for *in vivo* gene transfer into the adult mammalian brain: polyethylenimine: *Human Gene Therapy*, 7, 1947-1954, 1996.). The aliquots were administered by tail vein injection either as a single dose of 100 μ g or divided dose of 150 μ g and 200 μ g on consecutive days. Five days post-injection, liver tissue was removed for DNA and RNA isolation. DNA was isolated as previously described (Kren, B. T., Trembley, J. H. & Steer, C. J., "Alterations in mRNA stability during rat liver regeneration," *Am. J. Physiol.*, 270, G763-G777, 1996) for PCR amplification of exon 8 of the rat factor IX gene. RNA was isolated for RT-PCR amplification of the same region as the genomic DNA using RNAexol and RNAmate (Intermountain Scientific Corp., Kaysville, UT) according to the manufacturer's protocol.

Factor IX activity assay. Blood samples from vehicle (n = 9) and oligonucleotide-treated (n = 3) rats were collected 20 days after the second tail vein injection in 0.1 vol. of 0.105 M sodium citrate/citric acid. After centrifugation at 2,500 x g and then 15,000 x g the resulting plasma was stored at -70°C. The factor IX activity was determined from activated partial thromboplastin time (APTT) assays. Briefly, 50 μ l of APTT reagent (DADE, Miami, FL), 50 μ l of human factor IX-deficient plasma (George King Biomedical, Overland, KS), and 50 μ l of 1/10 dilution of rat test plasma in Hepes buffer (50 mM Hepes/100 mM NaCl/0.02% NaN₃, pH 7.4) were incubated at 37°C for 3 min in an ST4 coagulometer (American Bioproducts, Parsippany, NJ). Clotting was initiated by addition of 50 μ l of 33 mM CaCl₂ in Hepes buffer. Factor IX activity of duplicate samples was determined from a log-log standard curve constructed from the APTT results for dilution (1:10 to 1:80) of pooled plasma from normal male rats (n = 12).

DNA/RNA isolation and cloning. The cells were harvested by scrapping 48 h after transfection. Genomic DNA larger than 100-150 base pairs was isolated using the high

pure PCR template preparation kit (Boehringer Mannheim, Corp., Indianapolis, IN). RNA was isolated using RNAzol™ B (Tel-Test, Inc., Friendswood, TX), according to the manufacturer's protocol. PCR amplification of a 374-nt fragment of the rat factor IX gene was performed with 300 ng of the isolated DNA from either the primary hepatocytes or liver tissue. The primers were designed as 5'-ATTGCCTTGCTGGAACTGGATAAAC-3' (SEQ ID No. 48) and 5'TTGCCTTTCATTGCACATTCTTCAC-3' (SEQ ID No. 49) (Oligos Etc., Wilsonville, OR) corresponding to nucleotides 433-457 and 782-806, respectively, of the rat factor IX cDNA. Primers were annealed at 59°C for 20 sec, elongation was for 45 sec at 72°C and denaturation proceeded for 45 sec at 94°C. The sample was amplified for 30 cycles using Expand Hi-fidelity™ polymerase (Boehringer Mannheim, Corp.). The PCR amplification products from both the hepatocytes and intact liver factor IX genes were subcloned into the TA cloning vector pCR™2.1 (Invitrogen, San Diego, CA), and the ligated material used to transform frozen competent *E. coli*. To rule out PCR artifacts 300 ng of control DNA was incubated with 0.5, 1.0 and 1.5 µg of the oligonucleotide prior to the PCR-amplification reaction. Additionally, 1.0 µg of the chimeric alone was used as the "template" for the PCR amplification.

RT-PCR amplification was done utilizing the Titian™ one tube RT-PCR system (Boehringer Mannheim, Corp.) According to the manufacturer's protocol using the same primers as those used for the DNA PCR amplification. To rule out DNA contamination, the RNA samples were treated with RQ1 DNase free RNase (Promega Corp., Madison, WI) and RT-PCR negative controls of RNased RNA samples were performed in parallel with the RT-PCR reaction. Each of the PCR reactions were ligated into the same TA cloning vector and transformed into frozen competent *E. coli*.

Colony hybridization and sequencing. Eighteen to 20 h after plating, the colonies were lifted onto MSI MagnaGraph nylon filters, replicated and processed for hybridization according to the manufacturer's recommendation. The filters were hybridized for 24 h with 17 mer oligonucleotide probes 365-A (5'AAGGAGATAGTGGGGGA-3') (SEQ ID No. 50) OR 365-C (5'AAGGAGATCGTGGGGGA-3') (SEQ ID No. 51) (Life technologies, Inc., Gaithersburg, MD), where the underlined nucleotide is the target for

mutagenesis. The probes were ³²P-end-labeled using (γ-³²P) ATP (> 7,000 Ci/mmol) and T4 polynucleotide kinase (New England Biolabs, Inc., Beverly MA). Hybridizations were performed at 37°C in 2X sodium chloride sodium citrate containing 1% SDS, 5X Denhardt's and 200 μg/ml denatured sonicated fish sperm DNA. After hybridization, the filters were rinsed in 1X sodium chloride sodium phosphate EDTA, 0.5% SDS and then washed at 54°C for 1 h in 50 mM Tris-HCl, pH 8.0 containing 3 M tetramethylammonium chloride, 2 mM EDTA, pH 8.0, 0.1% SDS (Melchior, W. B. & Von Hippel, P. H. "Alteration of the relative stability of dA.dT and dG.dC base pairs in DNA," Proc. Natl. Acad. Sci. USA, 70, 298-302, 1973.). Autoradiography was performed with NEN®Reflection film at -70°C using an intensifying screen. Plasmid DNA was prepared from colonies identified as hybridizing with 365-A or 365-C using Qiagen miniprep kit (Chatsworth, CA) and subjected to automatic sequencing using the mp13 forward and reverse primers as well as a gene specific primer, 5'GTTGACCGAGCCACATGCCTTAG-3' (SEQ ID No. 52) corresponding to nucleotides 616 to 638 of the rat factor IX cDNA using an ABI 370A sequencer (Perkin-Elmer, Corp., Foster City, CA).

7.4 EXAMPLES OF CMV USEFUL FOR THE REDUCTION OF LDL LEVELS IN HUMANS

A CMV suitable for the modification of Apo B having a sequence comprising the sequence of SEQ ID No: 5 is given below.

Apo B 41/UR (mut→WT) (SEQ ID No. 53)

```

u GCGCG gac ccg acc gaa uuc ggu aac ugu au
u
u
u CGCGC CTG GGC TGG CTT AAG CCA TTG ACA Tu
      3' 5'

```

A CMV suitable for the modification of Apo B having a sequence comprising the sequence of SEQ ID No: 12 is given below.

Apo B 5/U88 (mut→WT)

(SEQ ID No. 54)

```

u GCGCG cug uuc aaa gug uaC GGA TCC ucu uug acu gac gau
u
u
u CGCGC3' GAC5' AAG TTT CAC ATG CCT AGG AGA AAC TGA CTG CTu

```

7.5 CORRECTION OF A CRIGLER-NAJJAR-LIKE MUTATION IN THE GUNN RAT

Mutant rats with hyperbilirubinemia, termed Gunn rats, have a single nucleotide deletion in the gene encoding bilirubin-uridinediphosphoglucuronate glucuronosyltransferase (*UGT1A1*). Roy Chowdhury, J., et al., 1991, J. Biol. Chem. **266**, 18294. Human patients with Crigler-Najjar syndrome type I also have mutations of the *UGT1A1* gene, resulting in life-long hyperbilirubinemia and consequent brain damage. Bosma, P.J., et al., 1992, FASEB J. **6**, 2859; Jansen, P.L.M., et al., Progress In Liver Diseases, XIII, Boyer, J.L., & Ockner, R.K., editors (W.B. Saunders, Phil. 1995), pp 125-150. The structure of CN3, a CMV designed to correct the Gunn rat mutation is given below.

CN3 (mut→WT)

(SEQ ID No. 55)

```

T GCGCG gg gac uua caG GAC CTT TAC uga ctt cua T
T
T
T CGCGC CC CTG AAT GTC CTG GAA ATG ACT GCC GAT T
3' 5'

```

Gunn rat primary cultured hepatocytes were treated with 150 nM CN3 according to the above protocol except that the carrier was either the negatively charged glycosylated lipid vesicles of section 6.2.2 or a lactosylated-PEI carrier at a ratio of

oligonucleotide phosphate to imine of 1:4 . The results were 8.5% conversion with the negatively charged liposome and 3.6% conversion with lactosylated-PEI carrier.

Gunn rats were injected with 1 mg/Kg of CN3 complexed with either 25 kDa Lac-PEI or complexed with negatively charged Gc lipid vesicles (Gc-NLV) as described above. The rate of gene conversion was determined by cloning and hybridization according to the procedure described for factor IX. The results shown below indicate that between about 15% and 25% of the copies of the *UCT1A1* gene were converted.

**Frequency of Insertion of G at nucleotide 1239 of the UGT-1 Gene
(In Gunn Rats)**

Vehicle	Dosage	G Clones/Total Clones	Frequency (%)
Gc-NLV	1 mg	112/815	15.4
		208/761	27.3
		185/974	18.9
		39/273	14.6 ¹
		78/403	19.3 ²
25 kDa PEI (Lactosylated)	1 mg	188/838	22.4
		254/1150	22.1
		245/997	24.6

¹Initial conversion frequency determined.

²Conversion frequency determined 7 days after 70% partial hepatectomy.

A Gunn rat was injected on five successive days with 1mg/Kg of CN3 complexed with 25 kDa Lac-PEI as above. Twenty five days after the final injection the serum bilirubin had declined from 6.2 mg/dl to 3.5 mg/dl and remained at that level for a further 25 days.

7.6 CORRECTION OF A FACTOR IX MUTATION IN DOG

The Chapel Hill strain of dogs, which has a (G→A)¹⁴⁷⁷ mutation that results in hemophilia in the animals, was used to obtain primary cultured hepatocytes. Four CMV to correct this mutation have been synthesized.

DIX1 (mut→WT) (SEQ ID No. 56)

```

T gcgcg auu caa aga aTT GAC CCT AAT AAT cga ccc cT
T
T
T CGCGC TAA GTT TCT TAA CTG GGA TTA TTA GCT GGG GT
      3' 5'

```

DIX2 (mut→WT) (SEQ ID No. 57)

```

T gcgcg caa aga auu gAC CCT AAT aaU cga cT
T
T
T CGCGC GTT TCT TAA CTG GGA TTA TTA GCT GT
      3' 5'

```

DIX3 (mut→WT) (SEQ ID No. 58)
 u gcgcg auu caa aga auu gac ccu aau aau cga ccc cu
 u
 u CGCGC TAA GTT TCT TAA CTG GGA TTA TTA GCT GGG Gu
 3' 5'

DIX4 (mut→WT) (SEQ ID No. 59)
 u gcgcg auu caa aga auu gac uccu aau aau cga ccc cu
 u
 u CGCGC TAA GTT TCT TAA CTG GGA TTA TTA GCT GGG Gu
 3' 5'

DIX1 differs from DIX3 by the replacement of the intervening DNA segment with 2'-O-methyl RNA and replacement of the tetrathymidine linkers with tetrauracil. DIX 4 differs from DIX3 in that the mutational vector contains a mismatch in the mutator region. In DIX4 the 5' (lower) strand encodes the desired (wild-type) sequence while the 3' (upper) strand has the sequence of the target, i.e., the mutant sequence.

The hepatocytes were treated with 360 nM DIX1 complexed in either 25 kDa Lac-PEI or galactocerebroside-containing aqueous-cored, negatively charged lipid vesicles (Gc-NLV). The results are given in the table below.

Frequency of conversion of A to G at nucleotide 1477 of the Factor IX Gene
(Primary Hepatocytes from the Chapel Hill Strain of Hemophilia B Dogs)

Vehicle	Number of Times Transfected	Concentration	G Clones/Total Clones	Frequency (%)
Gc-NLV	Once	360 nM	30/195	15.44
			30/218	13.76
	Twice		30/118	25.4
Lac-PEI 25 kDa	Once*	360 nM	20/141	14.2
			48/348	13.3
	Twice		21/107	19.6

*RT-PCR on parallel transfected cultures gave an A to G conversion frequency of 11.1%

Each of the DIX2-DIX4 were also tested on primary cultured dog hepatocytes as above. The results showed that DIX2 worked poorly, possibly due to the low (25%) GC percentage. The subsequent experiments the results of DIX3 were about 16% conversion, while a parallel experiments DIX1 gave 10% conversion and the results of DIX4 were about as good as DIX1.

GenBank Sequence References for the Exons of the Human Apolipoprotein B-100 Gene

TABLE II

Exon No.	cDNA Boundary	GenBank Accession No. Sequence Reference
1	126 to 207	M19808
2	208 to 246	M19808
3	247 to 362	M19809
4	363 to 508	M19810
5	509 to 662	M19811
6	663 to 818	M19812
7	819 to 943	M19813
8	944 to 1029	M19813
9	1030 to 1249	M19815
10	1250 to 1477	M19816
11	1478 to 1595	M19818
12	1596 to 1742	M19818
13	1743 to 1954	M19820
14	1955 to 2192	M19820
15	2193 to 2359	M19821
16	2360 to 2561	M19823
17	2562 to 2729	M19824
18	2730 to 2941	M19824
19	2942 to 3124	M19825
20	3125 to 3246	M19825
21	3247 to 3457	M19827
22	3458 to 3633	M19828
23	3634 to 3821	M19828
24	3822 to 3967	M19828
25	3968 to 4341	M19828
26	4342 to 11913	M19828
27	11914 to 12028	M19828
28	12029 to 12212	M19828
29	12213 to 13816	M19828

SEQ ID	TABLE I			G/C#	NA Change	AA Change	AA	%APOB100	Restriction Site
No.	Sequence (5'→3')								
4	ACTCTGGATGGG	I	AAGCCGCCCTCA	15	A→T	K→Stop	1701	36.9	None
5	CTGGGCTGGCTT	A	AGCCATTGACAT	13	C→A	S→TAA	1876	40.8	+CTTAAG
6	GCTCTCTGGGG	A	IACATACTGGGC	14	G→T	E→Stop	1921	41.8	None
7	GATGCCGTTGAG	I	AGCCCCAAGAAT	13	A→T	K→Stop	2047	44.5	None
8	GAGAGGAATCG	A	IAAACCATATAG	10	C→T	Q→Stop	2085	45.4	+ATCGAT
9	TGTAAGAAAATA	A	AGAGCAGCCCTG	10	C→A	Y→Stop	2110	45.9	None
10	GCAGCCCTGGG	A	IAACTCCACAGC	16	A→T	K→Stop	2116	46.0	None
11	GCAAGCTAATG	A	TAGCTGAATTCATTC	8	T→G	Y→Stop	2124	46.2	+AGCT
12	CAAGTTTCACAT	G	CGCAGGAGAACTG	11	A→T	K→Stop	2138	46.5	+CCTAGG
13	ATATACAAATT	G	CATGAGATGATGCC	9	T→G	L→Stop	2159	47.0	+CATG
14	AACTATCTCA	A	CTGAGACATATATG	8	C→T	Q→Stop	2174	47.3	-CTGCAG
15	GCTAATATTAT	G	TAAATCATTGAAATTA	3	G→T	E→Stop	2204	48.0	+TTAA
16	TCATGAGCACT	A	GCATATCCGTGTA	11	T→G	Y→Stop	2216	48.3	+CTAG
17	CTGCAGCAGCT	T	IAGAGACACATAC	12	A→T	K→Stop	2270	49.4	-CTTAAG
18	AACAGTGAGCT	G	IAGTGGCCCGTTC	14	C→T	Q→Stop	2684	58.6	None
19	CAGACTCCGTT	A	ACCAGAAATCGC	12	T→A	L→Stop	2712	59.2	+GTTAAC
20	AAAGGTCATGG	I	AATGGGCCTGCC	14	A→T	K→Stop	2930	64.0	None
21	ACATATATGAT	A	IAATTTGATCAGT	5	C→T	Q→Stop	2180	47.5	Physiologic
22	ATGGAGGACGT	G	IGCGGCCGCCCTGG	18	C→T	R→C	112	Apo E	None
23	GACCTGCAGAA	G	IGCCTGGCAGTGT	15	C→T	R→C	158	Apo E	None
24	GACCTGCAGAA	G	CGCCTGGCAGTGT	16	T→C	C→R	158	Apo E	None
25	TAAGGTCAGG	A	ITGAGACCCAGCC	13	A→T	NA	-491	Apo E	None

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: Steer, Clifford J.

Kren, Betsy T.

Bandyopadhyay, Paramita

Roy-Chowdhury, Jayanta

(ii) TITLE OF THE INVENTION: In Vivo Use of Recombinagenic
Oligonucleobases to Correct Genetic Lesions in
Hepatocytes

(iii) NUMBER OF SEQUENCES: 59

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Kimeragen, Inc.

(B) STREET: 300 Pheasant Run

(C) CITY: Newtown

(D) STATE: PA

(E) COUNTRY: USA

(F) ZIP: 18940

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette

(B) COMPUTER: IBM Compatible

(C) OPERATING SYSTEM: DOS

(D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE:

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 60/054,288

(B) FILING DATE: 30-APR-1997

(A) APPLICATION NUMBER: 60/054,837

(B) FILING DATE: 05-AUG-1997

(A) APPLICATION NUMBER: 60/064,996

(B) FILING DATE: 10-NOV-1997

(A) APPLICATION NUMBER: 60/074,497

(B) FILING DATE: 12-FEB-1998

(A) APPLICATION NUMBER: PCT US 98/08834

(B) FILING DATE: 30-APR-1998

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Hansburg, Daniel

(B) REGISTRATION NUMBER: 36156

(C) REFERENCE/DOCKET NUMBER: 7991-033-999

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 215-504-4444

(B) TELEFAX: 215-504-4545

(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4563 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Met	Asp	Pro	Pro	Arg	Pro	Ala	Leu	Leu	Ala	Leu	Leu	Ala	Leu	Pro	Ala
1				5					10					15	
Leu	Leu	Leu	Leu	Leu	Leu	Ala	Gly	Ala	Arg	Ala	Glu	Glu	Glu	Met	Leu
			20					25					30		
Glu	Asn	Val	Ser	Leu	Val	Cys	Pro	Lys	Asp	Ala	Thr	Arg	Phe	Lys	His
		35				40					45				
Leu	Arg	Lys	Tyr	Thr	Tyr	Asn	Tyr	Glu	Ala	Glu	Ser	Ser	Ser	Gly	Val
	50					55				60					
Pro	Gly	Thr	Ala	Asp	Ser	Arg	Ser	Ala	Thr	Arg	Ile	Asn	Cys	Lys	Val
65					70					75				80	
Glu	Leu	Glu	Val	Pro	Gln	Leu	Cys	Ser	Phe	Ile	Leu	Lys	Thr	Ser	Gln
			85					90					95		
Cys	Ile	Leu	Lys	Glu	Val	Tyr	Gly	Phe	Asn	Pro	Glu	Gly	Lys	Ala	Leu
		100						105					110		
Leu	Lys	Lys	Thr	Lys	Asn	Ser	Glu	Glu	Phe	Ala	Ala	Ala	Met	Ser	Arg
		115					120						125		
Tyr	Glu	Leu	Lys	Leu	Ala	Ile	Pro	Glu	Gly	Lys	Gln	Val	Phe	Leu	Tyr

130		135		140
Pro Glu Lys Asp Glu	Pro Thr Tyr Ile Leu	Asn Ile Lys Arg Gly Ile		
145	150	155	160	
Ile Ser Ala Leu Leu	Val Pro Pro Glu Thr	Glu Glu Ala Lys Gln Val		
	165	170	175	
Leu Phe Leu Asp Thr	Val Tyr Gly Asn Cys	Ser Thr His Phe Thr Val		
	180	185	190	
Lys Thr Arg Lys Gly	Asn Val Ala Thr Glu	Ile Ser Thr Glu Arg Asp		
	195	200	205	
Leu Gly Gln Cys Asp	Arg Phe Lys Pro Ile	Arg Thr Gly Ile Ser Pro		
	210	215	220	
Leu Ala Leu Ile Lys	Gly Met Thr Arg Pro	Leu Ser Thr Leu Ile Ser		
225	230	235	240	
Ser Ser Gln Ser Cys	Gln Tyr Thr Leu Asp	Ala Lys Arg Lys His Val		
	245	250	255	
Ala Glu Ala Ile Cys	Lys Glu Gln His Leu	Phe Leu Pro Phe Ser Tyr		
	260	265	270	
Lys Asn Lys Tyr Gly	Met Val Ala Gln Val	Thr Gln Thr Leu Lys Leu		
	275	280	285	
Glu Asp Thr Pro Lys	Ile Asn Ser Arg Phe	Phe Gly Glu Gly Thr Lys		
	290	295	300	
Lys Met Gly Leu Ala	Phe Glu Ser Thr Lys	Ser Thr Ser Pro Pro Lys		
305	310	315	320	
Gln Ala Glu Ala Val	Leu Lys Thr Val Gln	Glu Leu Lys Lys Leu Thr		
	325	330	335	
Ile Ser Glu Gln Asn	Ile Gln Arg Ala Asn	Leu Phe Asn Lys Leu Val		
	340	345	350	
Thr Glu Leu Arg Gly	Leu Ser Asp Glu Ala	Val Thr Ser Leu Leu Pro		
	355	360	365	
Gln Leu Ile Glu Val	Ser Ser Pro Ile Thr	Leu Gln Ala Leu Val Gln		
	370	375	380	
Cys Gly Gln Pro Gln	Cys Ser Thr His Ile	Leu Gln Trp Leu Lys Arg		
385	390	395	400	
Val His Ala Asn Pro	Leu Leu Ile Asp Val	Val Thr Tyr Leu Val Ala		
	405	410	415	
Leu Ile Pro Glu Pro	Ser Ala Gln Gln Leu	Arg Glu Ile Phe Asn Met		
	420	425	430	
Ala Arg Asp Gln Arg	Ser Arg Ala Thr Leu	Tyr Ala Leu Ser His Ala		
	435	440	445	
Val Asn Asn Tyr His	Lys Thr Asn Pro Thr	Gly Thr Gln Glu Leu Leu		
	450	455	460	
Asp Ile Ala Asn Tyr	Leu Met Glu Gln Ile	Gln Asp Asp Cys Thr Gly		
465	470	475	480	
Asp Glu Asp Tyr Thr	Tyr Leu Ile Leu Arg	Val Ile Gly Asn Met Gly		
	485	490	495	

Gln Thr Met Glu Gln Leu Thr Pro Glu Leu Lys Ser Ser Ile Leu Lys			
500	505	510	
Cys Val Gln Ser Thr Lys Pro Ser Leu Met Ile Gln Lys Ala Ala Ile			
515	520	525	
Gln Ala Leu Arg Lys Met Glu Pro Lys Asp Lys Asp Gln Glu Val Leu			
530	535	540	
Leu Gln Thr Phe Leu Asp Asp Ala Ser Pro Gly Asp Lys Arg Leu Ala			
545	550	555	560
Ala Tyr Leu Met Leu Met Arg Ser Pro Ser Gln Ala Asp Ile Asn Lys			
565	570	575	
Ile Val Gln Ile Leu Pro Trp Glu Gln Asn Glu Gln Val Lys Asn Phe			
580	585	590	
Val Ala Ser His Ile Ala Asn Ile Leu Asn Ser Glu Glu Leu Asp Ile			
595	600	605	
Gln Asp Leu Lys Lys Leu Val Lys Glu Val Leu Lys Glu Ser Gln Leu			
610	615	620	
Pro Thr Val Met Asp Phe Arg Lys Phe Ser Arg Asn Tyr Gln Leu Tyr			
625	630	635	640
Lys Ser Val Ser Ile Pro Ser Leu Asp Pro Ala Ser Ala Lys Ile Glu			
645	650	655	
Gly Asn Leu Ile Phe Asp Pro Asn Asn Tyr Leu Pro Lys Glu Ser Met			
660	665	670	
Leu Lys Thr Thr Leu Thr Ala Phe Gly Phe Ala Ser Ala Asp Leu Ile			
675	680	685	
Glu Ile Gly Leu Glu Gly Lys Gly Phe Glu Pro Thr Leu Glu Ala Leu			
690	695	700	
Phe Gly Lys Gln Gly Phe Phe Pro Asp Ser Val Asn Lys Ala Leu Tyr			
705	710	715	720
Trp Val Asn Gly Gln Val Pro Asp Gly Val Ser Lys Val Leu Val Asp			
725	730	735	
His Phe Gly Tyr Thr Lys Asp Asp Lys His Glu Gln Asp Met Val Asn			
740	745	750	
Gly Ile Met Leu Ser Val Glu Lys Leu Ile Lys Asp Leu Lys Ser Lys			
755	760	765	
Glu Val Pro Glu Ala Arg Ala Tyr Leu Arg Ile Leu Gly Glu Glu Leu			
770	775	780	
Gly Phe Ala Ser Leu His Asp Leu Gln Leu Leu Gly Lys Leu Leu Leu			
785	790	795	800
Met Gly Ala Arg Thr Leu Gln Gly Ile Pro Gln Met Ile Gly Glu Val			
805	810	815	
Ile Arg Lys Gly Ser Lys Asn Asp Phe Phe Leu His Tyr Ile Phe Met			
820	825	830	
Glu Asn Ala Phe Glu Leu Pro Thr Gly Ala Gly Leu Gln Leu Gln Ile			
835	840	845	
Ser Ser Ser Gly Val Ile Ala Pro Gly Ala Lys Ala Gly Val Lys Leu			

850	855	860
Glu Val Ala Asn Met	Gln Ala Glu Leu Val	Ala Lys Pro Ser Val Ser
865	870	875
Val Glu Phe Val Thr	Asn Met Gly Ile Ile	Ile Pro Asp Phe Ala Arg
885	890	895
Ser Gly Val Gln Met	Asn Thr Asn Phe Phe	His Glu Ser Gly Leu Glu
900	905	910
Ala His Val Ala Leu	Lys Pro Gly Lys Leu	Lys Phe Ile Ile Pro Ser
915	920	925
Pro Lys Arg Pro Val	Lys Leu Leu Ser Gly	Gly Asn Thr Leu His Leu
930	935	940
Val Ser Thr Thr Lys	Thr Glu Val Ile Pro	Pro Leu Ile Glu Asn Arg
945	950	955
Gln Ser Trp Ser Val	Cys Lys Gln Val Phe	Pro Gly Leu Asn Tyr Cys
965	970	975
Thr Ser Gly Ala Tyr	Ser Asn Ala Ser Ser	Thr Asp Ser Ala Ser Tyr
980	985	990
Tyr Pro Leu Thr Gly	Asp Thr Arg Leu Glu	Leu Glu Leu Arg Pro Thr
995	1000	1005
Gly Glu Ile Glu Gln	Tyr Ser Val Ser Ala	Thr Tyr Glu Leu Gln Arg
1010	1015	1020
Glu Asp Arg Ala Leu	Val Asp Thr Leu Lys	Phe Val Thr Gln Ala Glu
1025	1030	1035
Gly Ala Lys Gln Thr	Glu Ala Thr Met Thr	Phe Lys Tyr Asn Arg Gln
1045	1050	1055
Ser Met Thr Leu Ser	Ser Glu Val Gln Ile	Pro Asp Phe Asp Val Asp
1060	1065	1070
Leu Gly Thr Ile Leu	Arg Val Asn Asp Glu	Ser Thr Glu Gly Lys Thr
1075	1080	1085
Ser Tyr Arg Leu Thr	Leu Asp Ile Gln Asn	Lys Lys Ile Thr Glu Val
1090	1095	1100
Ala Leu Met Gly His	Leu Ser Cys Asp Thr	Lys Glu Glu Arg Lys Ile
1105	1110	1115
Lys Gly Val Ile Ser	Ile Pro Arg Leu Gln	Ala Glu Ala Arg Ser Glu
1125	1130	1135
Ile Leu Ala His Trp	Ser Pro Ala Lys Leu	Leu Leu Gln Met Asp Ser
1140	1145	1150
Ser Ala Thr Ala Tyr	Gly Ser Thr Val Ser	Lys Arg Val Ala Trp His
1155	1160	1165
Tyr Asp Glu Glu Lys	Ile Glu Phe Glu Trp	Asn Thr Gly Thr Asn Val
1170	1175	1180
Asp Thr Lys Lys Met	Thr Ser Asn Phe Pro	Val Asp Leu Ser Asp Tyr
1185	1190	1195
Pro Lys Ser Leu His	Met Tyr Ala Asn Arg	Leu Leu Asp His Arg Val
1205	1210	1215

Pro Gln Thr Asp Met Thr Phe Arg His Val Gly Ser Lys Leu Ile Val
 1220 1225 1230
 Ala Met Ser Ser Trp Leu Gln Lys Ala Ser Gly Ser Leu Pro Tyr Thr
 1235 1240 1245
 Gln Thr Leu Gln Asp His Leu Asn Ser Leu Lys Glu Phe Asn Leu Gln
 1250 1255 1260
 Asn Met Gly Leu Pro Asp Phe His Ile Pro Glu Asn Leu Phe Leu Lys
 265 1270 1275 1280
 Ser Asp Gly Arg Val Lys Tyr Thr Leu Asn Lys Asn Ser Leu Lys Ile
 1285 1290 1295
 Glu Ile Pro Leu Pro Phe Gly Gly Lys Ser Ser Arg Asp Leu Lys Met
 1300 1305 1310
 Leu Glu Thr Val Arg Thr Pro Ala Leu His Phe Lys Ser Val Gly Phe
 1315 1320 1325
 His Leu Pro Ser Arg Glu Phe Gln Val Pro Thr Phe Thr Ile Pro Lys
 1330 1335 1340
 Leu Tyr Gln Leu Gln Val Pro Leu Leu Gly Val Leu Asp Leu Ser Thr
 345 1350 1355 1360
 Asn Val Tyr Ser Asn Leu Tyr Asn Trp Ser Ala Ser Tyr Ser Gly Gly
 1365 1370 1375
 Asn Thr Ser Thr Asp His Phe Ser Leu Arg Ala Arg Tyr His Met Lys
 1380 1385 1390
 Ala Asp Ser Val Val Asp Leu Leu Ser Tyr Asn Val Gln Gly Ser Gly
 1395 1400 1405
 Glu Thr Thr Tyr Asp His Lys Asn Thr Phe Thr Leu Ser Cys Asp Gly
 1410 1415 1420
 Ser Leu Arg His Lys Phe Leu Asp Ser Asn Ile Lys Phe Ser His Val
 425 1430 1435 1440
 Glu Lys Leu Gly Asn Asn Pro Val Ser Lys Gly Leu Leu Ile Phe Asp
 1445 1450 1455
 Ala Ser Ser Ser Trp Gly Pro Gln Met Ser Ala Ser Val His Leu Asp
 1460 1465 1470
 Ser Lys Lys Lys Gln His Leu Phe Val Lys Glu Val Lys Ile Asp Gly
 1475 1480 1485
 Gln Phe Arg Val Ser Ser Phe Tyr Ala Lys Gly Thr Tyr Gly Leu Ser
 1490 1495 1500
 Cys Gln Arg Asp Pro Asn Thr Gly Arg Leu Asn Gly Glu Ser Asn Leu
 505 1510 1515 1520
 Arg Phe Asn Ser Ser Tyr Leu Gln Gly Thr Asn Gln Ile Thr Gly Arg
 1525 1530 1535
 Tyr Glu Asp Gly Thr Leu Ser Leu Thr Ser Thr Ser Asp Leu Gln Ser
 1540 1545 1550
 Gly Ile Ile Lys Asn Thr Ala Ser Leu Lys Tyr Glu Asn Tyr Glu Leu
 1555 1560 1565
 Thr Leu Lys Ser Asp Thr Asn Gly Lys Tyr Lys Asn Phe Ala Thr Ser

1570	1575	1580
Asn Lys Met Asp Met Thr Phe Ser Lys Gln Asn Ala Leu Leu Arg Ser		
585	1590	1595 1600
Glu Tyr Gln Ala Asp Tyr Glu Ser Leu Arg Phe Phe Ser Leu Leu Ser		
1605	1610	1615
Gly Ser Leu Asn Ser His Gly Leu Glu Leu Asn Ala Asp Ile Leu Gly		
1620	1625	1630
Thr Asp Lys Ile Asn Ser Gly Ala His Lys Ala Thr Leu Arg Ile Gly		
1635	1640	1645
Gln Asp Gly Ile Ser Thr Ser Ala Thr Thr Asn Leu Lys Cys Ser Leu		
1650	1655	1660
Leu Val Leu Glu Asn Glu Leu Asn Ala Glu Leu Gly Leu Ser Gly Ala		
665	1670	1675 1680
Ser Met Lys Leu Thr Thr Asn Gly Arg Phe Arg Glu His Asn Ala Lys		
1685	1690	1695
Phe Ser Leu Asp Gly Lys Ala Ala Leu Thr Glu Leu Ser Leu Gly Ser		
1700	1705	1710
Ala Tyr Gln Ala Met Ile Leu Gly Val Asp Ser Lys Asn Ile Phe Asn		
1715	1720	1725
Phe Lys Val Ser Gln Glu Gly Leu Lys Leu Ser Asn Asp Met Met Gly		
1730	1735	1740
Ser Tyr Ala Glu Met Lys Phe Asp His Thr Asn Ser Leu Asn Ile Ala		
745	1750	1755 1760
Gly Leu Ser Leu Asp Phe Ser Ser Lys Leu Asp Asn Ile Tyr Ser Ser		
1765	1770	1775
Asp Lys Phe Tyr Lys Gln Thr Val Asn Leu Gln Leu Gln Pro Tyr Ser		
1780	1785	1790
Leu Val Thr Thr Leu Asn Ser Asp Leu Lys Tyr Asn Ala Leu Asp Leu		
1795	1800	1805
Thr Asn Asn Gly Lys Leu Arg Leu Glu Pro Leu Lys Leu His Val Ala		
1810	1815	1820
Gly Asn Leu Lys Gly Ala Tyr Gln Asn Asn Glu Ile Lys His Ile Tyr		
825	1830	1835 1840
Ala Ile Ser Ser Ala Ala Leu Ser Ala Ser Tyr Lys Ala Asp Thr Val		
1845	1850	1855
Ala Lys Val Gln Gly Val Glu Phe Ser His Arg Leu Asn Thr Asp Ile		
1860	1865	1870
Ala Gly Leu Ala Ser Ala Ile Asp Met Ser Thr Asn Tyr Asn Ser Asp		
1875	1880	1885
Ser Leu His Phe Ser Asn Val Phe Arg Ser Val Met Ala Pro Phe Thr		
1890	1895	1900
Met Thr Ile Asp Ala His Thr Asn Gly Asn Gly Lys Leu Ala Leu Trp		
905	1910	1915 1920
Gly Glu His Thr Gly Gln Leu Tyr Ser Lys Phe Leu Leu Lys Ala Glu		
1925	1930	1935

Pro Leu Ala Phe Thr Phe Ser His Asp Tyr Lys Gly Ser Thr Ser His
 1940 1945 1950
 His Leu Val Ser Arg Lys Ser Ile Ser Ala Ala Leu Glu His Lys Val
 1955 1960 1965
 Ser Ala Leu Leu Thr Pro Ala Glu Gln Thr Gly Thr Trp Lys Leu Lys
 1970 1975 1980
 Thr Gln Phe Asn Asn Asn Glu Tyr Ser Gln Asp Leu Asp Ala Tyr Asn
 985 1990 1995 2000
 Thr Lys Asp Lys Ile Gly Val Glu Leu Thr Gly Arg Thr Leu Ala Asp
 2005 2010 2015
 Leu Thr Leu Leu Asp Ser Pro Ile Lys Val Pro Leu Leu Leu Ser Glu
 2020 2025 2030
 Pro Ile Asn Ile Ile Asp Ala Leu Glu Met Arg Asp Ala Val Glu Lys
 2035 2040 2045
 Pro Gln Glu Phe Thr Ile Val Ala Phe Val Lys Tyr Asp Lys Asn Gln
 2050 2055 2060
 Asp Val His Ser Ile Asn Leu Pro Phe Phe Glu Thr Leu Gln Glu Tyr
 065 2070 2075 2080
 Phe Glu Arg Asn Arg Gln Thr Ile Ile Val Val Leu Glu Asn Val Gln
 2085 2090 2095
 Arg Asn Leu Lys His Ile Asn Ile Asp Gln Phe Val Arg Lys Tyr Arg
 2100 2105 2110
 Ala Ala Leu Gly Lys Leu Pro Gln Gln Ala Asn Asp Tyr Leu Asn Ser
 2115 2120 2125
 Phe Asn Trp Glu Arg Gln Val Ser His Ala Lys Glu Lys Leu Thr Ala
 2130 2135 2140
 Leu Thr Lys Lys Tyr Arg Ile Thr Glu Asn Asp Ile Gln Ile Ala Leu
 145 2150 2155 2160
 Asp Asp Ala Lys Ile Asn Phe Asn Glu Lys Leu Ser Gln Leu Gln Thr
 2165 2170 2175
 Tyr Met Ile Gln Phe Asp Gln Tyr Ile Lys Asp Ser Tyr Asp Leu His
 2180 2185 2190
 Asp Leu Lys Ile Ala Ile Ala Asn Ile Ile Asp Glu Ile Ile Glu Lys
 2195 2200 2205
 Leu Lys Ser Leu Asp Glu His Tyr His Ile Arg Val Asn Leu Val Lys
 2210 2215 2220
 Thr Ile His Asp Leu His Leu Phe Ile Glu Asn Ile Asp Phe Asn Lys
 225 2230 2235 2240
 Ser Gly Ser Ser Thr Ala Ser Trp Ile Gln Asn Val Asp Thr Lys Tyr
 2245 2250 2255
 Gln Ile Arg Ile Gln Ile Gln Glu Lys Leu Gln Gln Leu Lys Arg His
 2260 2265 2270
 Ile Gln Asn Ile Asp Ile Gln His Leu Ala Gly Lys Leu Lys Gln His
 2275 2280 2285
 Ile Glu Ala Ile Asp Val Arg Val Leu Leu Asp Gln Leu Gly Thr Thr

2290	2295	2300
Ile Ser Phe Glu Arg Ile Asn Asp Val Leu Glu His Val Lys His Phe		
305	2310	2315 2320
Val Ile Asn Leu Ile Gly Asp Phe Glu Val Ala Glu Lys Ile Asn Ala		
2325	2330	2335
Phe Arg Ala Lys Val His Glu Leu Ile Glu Arg Tyr Glu Val Asp Gln		
2340	2345	2350
Gln Ile Gln Val Leu Met Asp Lys Leu Val Glu Leu Ala His Gln Tyr		
2355	2360	2365
Lys Leu Lys Glu Thr Ile Gln Lys Leu Ser Asn Val Leu Gln Gln Val		
2370	2375	2380
Lys Ile Lys Asp Tyr Phe Glu Lys Leu Val Gly Phe Ile Asp Asp Ala		
385	2390	2395 2400
Val Lys Lys Leu Asn Glu Leu Ser Phe Lys Thr Phe Ile Glu Asp Val		
2405	2410	2415
Asn Lys Phe Leu Asp Met Leu Ile Lys Lys Leu Lys Ser Phe Asp Tyr		
2420	2425	2430
His Gln Phe Val Asp Glu Thr Asn Asp Lys Ile Arg Glu Val Thr Gln		
2435	2440	2445
Arg Leu Asn Gly Glu Ile Gln Ala Leu Glu Leu Pro Gln Lys Ala Glu		
2450	2455	2460
Ala Leu Lys Leu Phe Leu Glu Glu Thr Lys Ala Thr Val Ala Val Tyr		
465	2470	2475 2480
Leu Glu Ser Leu Gln Asp Thr Lys Ile Thr Leu Ile Ile Asn Trp Leu		
2485	2490	2495
Gln Glu Ala Leu Ser Ser Ala Ser Leu Ala His Met Lys Ala Lys Phe		
2500	2505	2510
Arg Glu Thr Leu Glu Asp Thr Arg Asp Arg Met Tyr Gln Met Asp Ile		
2515	2520	2525
Gln Gln Glu Leu Gln Arg Tyr Leu Ser Leu Val Gly Gln Val Tyr Ser		
2530	2535	2540
Thr Leu Val Thr Tyr Ile Ser Asp Trp Trp Thr Leu Ala Ala Lys Asn		
545	2550	2555 2560
Leu Thr Asp Phe Ala Glu Gln Tyr Ser Ile Gln Asp Trp Ala Lys Arg		
2565	2570	2575
Met Lys Ala Leu Val Glu Gln Gly Phe Thr Val Pro Glu Ile Lys Thr		
2580	2585	2590
Ile Leu Gly Thr Met Pro Ala Phe Glu Val Ser Leu Gln Ala Leu Gln		
2595	2600	2605
Lys Ala Thr Phe Gln Thr Pro Asp Phe Ile Val Pro Leu Thr Asp Leu		
2610	2615	2620
Arg Ile Pro Ser Val Gln Ile Asn Phe Lys Asp Leu Lys Asn Ile Lys		
625	2630	2635 2640
Ile Pro Ser Arg Phe Ser Thr Pro Glu Phe Thr Ile Leu Asn Thr Phe		
2645	2650	2655

His Ile Pro Ser Phe Thr Ile Asp Phe Val Glu Met Lys Val Lys Ile
 2660 2665 2670
 Ile Arg Thr Ile Asp Gln Met Leu Asn Ser Glu Leu Gln Trp Pro Val
 2675 2680 2685
 Pro Asp Ile Tyr Leu Arg Asp Leu Lys Val Glu Asp Ile Pro Leu Ala
 2690 2695 2700
 Arg Ile Thr Leu Pro Asp Phe Arg Leu Pro Glu Ile Ala Ile Pro Glu
 705 2710 2715 2720
 Phe Ile Ile Pro Thr Leu Asn Leu Asn Asp Phe Gln Val Pro Asp Leu
 2725 2730 2735
 His Ile Pro Glu Phe Gln Leu Pro His Ile Ser His Thr Ile Glu Val
 2740 2745 2750
 Pro Thr Phe Gly Lys Leu Tyr Ser Ile Leu Lys Ile Gln Ser Pro Leu
 2755 2760 2765
 Phe Thr Leu Asp Ala Asn Ala Asp Ile Gly Asn Gly Thr Thr Ser Ala
 2770 2775 2780
 Asn Glu Ala Gly Ile Ala Ala Ser Ile Thr Ala Lys Gly Glu Ser Lys
 785 2790 2795 2800
 Leu Glu Val Leu Asn Phe Asp Phe Gln Ala Asn Ala Gln Leu Ser Asn
 2805 2810 2815
 Pro Lys Ile Asn Pro Leu Ala Leu Lys Glu Ser Val Lys Phe Ser Ser
 2820 2825 2830
 Lys Tyr Leu Arg Thr Glu His Gly Ser Glu Met Leu Phe Phe Gly Asn
 2835 2840 2845
 Ala Ile Glu Gly Lys Ser Asn Thr Val Ala Ser Leu His Thr Glu Lys
 2850 2855 2860
 Asn Thr Leu Glu Leu Ser Asn Gly Val Ile Val Lys Ile Asn Asn Gln
 865 2870 2875 2880
 Leu Thr Leu Asp Ser Asn Thr Lys Tyr Phe His Lys Leu Asn Ile Pro
 2885 2890 2895
 Lys Leu Asp Phe Ser Ser Gln Ala Asp Leu Arg Asn Glu Ile Lys Thr
 2900 2905 2910
 Leu Leu Lys Ala Gly His Ile Ala Trp Thr Ser Ser Gly Lys Gly Ser
 2915 2920 2925
 Trp Lys Trp Ala Cys Pro Arg Phe Ser Asp Glu Gly Thr His Glu Ser
 2930 2935 2940
 Gln Ile Ser Phe Thr Ile Glu Gly Pro Leu Thr Ser Phe Gly Leu Ser
 945 2950 2955 2960
 Asn Lys Ile Asn Ser Lys His Leu Arg Val Asn Gln Asn Leu Val Tyr
 2965 2970 2975
 Glu Ser Gly Ser Leu Asn Phe Ser Lys Leu Glu Ile Gln Ser Gln Val
 2980 2985 2990
 Asp Ser Gln His Val Gly His Ser Val Leu Thr Ala Lys Gly Met Ala
 2995 3000 3005
 Leu Phe Gly Glu Gly Lys Ala Glu Phe Thr Gly Arg His Asp Ala His

3010	3015	3020
Leu Asn Gly Lys Val Ile Gly Thr Leu Lys Asn Ser Leu Phe Phe Ser		
025	3030	3040
Ala Gln Pro Phe Glu Ile Thr Ala Ser Thr Asn Asn Glu Gly Asn Leu		
3045	3050	3055
Lys Val Arg Phe Pro Leu Arg Leu Thr Gly Lys Ile Asp Phe Leu Asn		
3060	3065	3070
Asn Tyr Ala Leu Phe Leu Ser Pro Ser Ala Gln Gln Ala Ser Trp Gln		
3075	3080	3085
Val Ser Ala Arg Phe Asn Gln Tyr Lys Tyr Asn Gln Asn Phe Ser Ala		
3090	3095	3100
Gly Asn Asn Glu Asn Ile Met Glu Ala His Val Gly Ile Asn Gly Glu		
105	3110	3120
Ala Asn Leu Asp Phe Leu Asn Ile Pro Leu Thr Ile Pro Glu Met Arg		
3125	3130	3135
Leu Pro Tyr Thr Ile Ile Thr Thr Pro Pro Leu Lys Asp Phe Ser Leu		
3140	3145	3150
Trp Glu Lys Thr Gly Leu Lys Glu Phe Leu Lys Thr Thr Lys Gln Ser		
3155	3160	3165
Phe Asp Leu Ser Val Lys Ala Gln Tyr Lys Lys Asn Lys His Arg His		
3170	3175	3180
Ser Ile Thr Asn Pro Leu Ala Val Leu Cys Glu Phe Ile Ser Gln Ser		
185	3190	3200
Ile Lys Ser Phe Asp Arg His Phe Glu Lys Asn Arg Asn Asn Ala Leu		
3205	3210	3215
Asp Phe Val Thr Lys Ser Tyr Asn Glu Thr Lys Ile Lys Phe Asp Lys		
3220	3225	3230
Tyr Lys Ala Glu Lys Ser His Asp Glu Leu Pro Arg Thr Phe Gln Ile		
3235	3240	3245
Pro Gly Tyr Thr Val Pro Val Val Asn Val Glu Val Ser Pro Phe Thr		
3250	3255	3260
Ile Glu Met Ser Ala Phe Gly Tyr Val Phe Pro Lys Ala Val Ser Met		
265	3270	3280
Pro Ser Phe Ser Ile Leu Gly Ser Asp Val Arg Val Pro Ser Tyr Thr		
3285	3290	3295
Leu Ile Leu Pro Ser Leu Glu Leu Pro Val Leu His Val Pro Arg Asn		
3300	3305	3310
Leu Lys Leu Ser Leu Pro Asp Phe Lys Glu Leu Cys Thr Ile Ser His		
3315	3320	3325
Ile Phe Ile Pro Ala Met Gly Asn Ile Thr Tyr Asp Phe Ser Phe Lys		
3330	3335	3340
Ser Ser Val Ile Thr Leu Asn Thr Asn Ala Glu Leu Phe Asn Gln Ser		
345	3350	3360
Asp Ile Val Ala His Leu Leu Ser Ser Ser Ser Ser Val Ile Asp Ala		
3365	3370	3375

Leu Gln Tyr Lys Leu Glu Gly Thr Thr Arg Leu Thr Arg Lys Arg Gly
 3380 3385 3390
 Leu Lys Leu Ala Thr Ala Leu Ser Leu Ser Asn Lys Phe Val Glu Gly
 3395 3400 3405
 Ser His Asn Ser Thr Val Ser Leu Thr Thr Lys Asn Met Glu Val Ser
 3410 3415 3420
 Val Ala Thr Thr Thr Lys Ala Gln Ile Pro Ile Leu Arg Met Asn Phe
 425 3430 3435 3440
 Lys Gln Glu Leu Asn Gly Asn Thr Lys Ser Lys Pro Thr Val Ser Ser
 3445 3450 3455
 Ser Met Glu Phe Lys Tyr Asp Phe Asn Ser Ser Met Leu Tyr Ser Thr
 3460 3465 3470
 Ala Lys Gly Ala Val Asp His Lys Leu Ser Leu Glu Ser Leu Thr Ser
 3475 3480 3485
 Tyr Phe Ser Ile Glu Ser Ser Thr Lys Gly Asp Val Lys Gly Ser Val
 3490 3495 3500
 Leu Ser Arg Glu Tyr Ser Gly Thr Ile Ala Ser Glu Ala Asn Thr Tyr
 505 3510 3515 3520
 Leu Asn Ser Lys Ser Thr Arg Ser Ser Val Lys Leu Gln Gly Thr Ser
 3525 3530 3535
 Lys Ile Asp Asp Ile Trp Asn Leu Glu Val Lys Glu Asn Phe Ala Gly
 3540 3545 3550
 Glu Ala Thr Leu Gln Arg Ile Tyr Ser Leu Trp Glu His Ser Thr Lys
 3555 3560 3565
 Asn His Leu Gln Leu Glu Gly Leu Phe Phe Thr Asn Gly Glu His Thr
 3570 3575 3580
 Ser Lys Ala Thr Leu Glu Leu Ser Pro Trp Gln Met Ser Ala Leu Val
 585 3590 3595 3600
 Gln Val His Ala Ser Gln Pro Ser Ser Phe His Asp Phe Pro Asp Leu
 3605 3610 3615
 Gly Gln Glu Val Ala Leu Asn Ala Asn Thr Lys Asn Gln Lys Ile Arg
 3620 3625 3630
 Trp Lys Asn Glu Val Arg Ile His Ser Gly Ser Phe Gln Ser Gln Val
 3635 3640 3645
 Glu Leu Ser Asn Asp Gln Glu Lys Ala His Leu Asp Ile Ala Gly Ser
 3650 3655 3660
 Leu Glu Gly His Leu Arg Phe Leu Lys Asn Ile Ile Leu Pro Val Tyr
 665 3670 3675 3680
 Asp Lys Ser Leu Trp Asp Phe Leu Lys Leu Asp Val Thr Thr Ser Ile
 3685 3690 3695
 Gly Arg Arg Gln His Leu Arg Val Ser Thr Ala Phe Val Tyr Thr Lys
 3700 3705 3710
 Asn Pro Asn Gly Tyr Ser Phe Ser Ile Pro Val Lys Val Leu Ala Asp
 3715 3720 3725
 Lys Phe Ile Ile Pro Gly Leu Lys Leu Asn Asp Leu Asn Ser Val Leu

3730	3735	3740
Val Met Pro Thr Phe His Val Pro Phe Thr Asp Leu Gln Val Pro Ser		
745	3750	3755 3760
Cys Lys Leu Asp Phe Arg Glu Ile Gln Ile Tyr Lys Lys Leu Arg Thr		
3765	3770	3775
Ser Ser Phe Ala Leu Asn Leu Pro Thr Leu Pro Glu Val Lys Phe Pro		
3780	3785	3790
Glu Val Asp Val Leu Thr Lys Tyr Ser Gln Pro Glu Asp Ser Leu Ile		
3795	3800	3805
Pro Phe Phe Glu Ile Thr Val Pro Glu Ser Gln Leu Thr Val Ser Gln		
3810	3815	3820
Phe Thr Leu Pro Lys Ser Val Ser Asp Gly Ile Ala Ala Leu Asp Leu		
825	3830	3835 3840
Asn Ala Val Ala Asn Lys Ile Ala Asp Phe Glu Leu Pro Thr Ile Ile		
3845	3850	3855
Val Pro Glu Gln Thr Ile Glu Ile Pro Ser Ile Lys Phe Ser Val Pro		
3860	3865	3870
Ala Gly Ile Ala Ile Pro Ser Phe Gln Ala Leu Thr Ala Arg Phe Glu		
3875	3880	3885
Val Asp Ser Pro Val Tyr Asn Ala Thr Trp Ser Ala Ser Leu Lys Asn		
3890	3895	3900
Lys Ala Asp Tyr Val Glu Thr Val Leu Asp Ser Thr Cys Ser Ser Thr		
905	3910	3915 3920
Val Gln Phe Leu Glu Tyr Glu Leu Asn Val Leu Gly Thr His Lys Ile		
3925	3930	3935
Glu Asp Gly Thr Leu Ala Ser Lys Thr Lys Gly Thr Phe Ala His Arg		
3940	3945	3950
Asp Phe Ser Ala Glu Tyr Glu Glu Asp Gly Lys Tyr Glu Gly Leu Gln		
3955	3960	3965
Glu Trp Glu Gly Lys Ala His Leu Asn Ile Lys Ser Pro Ala Phe Thr		
3970	3975	3980
Asp Leu His Leu Arg Tyr Gln Lys Asp Lys Lys Gly Ile Ser Thr Ser		
985	3990	3995 4000
Ala Ala Ser Pro Ala Val Gly Thr Val Gly Met Asp Met Asp Glu Asp		
4005	4010	4015
Asp Asp Phe Ser Lys Trp Asn Phe Tyr Tyr Ser Pro Gln Ser Ser Pro		
4020	4025	4030
Asp Lys Lys Leu Thr Ile Phe Lys Thr Glu Leu Arg Val Arg Glu Ser		
4035	4040	4045
Asp Glu Glu Thr Gln Ile Lys Val Asn Trp Glu Glu Glu Ala Ala Ser		
4050	4055	4060
Gly Leu Leu Thr Ser Leu Lys Asp Asn Val Pro Lys Ala Thr Gly Val		
065	4070	4075 4080
Leu Tyr Asp Tyr Val Asn Lys Tyr His Trp Glu His Thr Gly Leu Thr		
4085	4090	4095

Leu Arg Glu Val Ser Ser Lys Leu Arg Arg Asn Leu Gln Asp His Ala
 4100 4105 4110
 Glu Trp Val Tyr Gln Gly Ala Ile Arg Glu Ile Asp Asp Ile Asp Glu
 4115 4120 4125
 Arg Phe Gln Lys Gly Ala Ser Gly Thr Thr Gly Thr Tyr Gln Glu Trp
 4130 4135 4140
 Lys Asp Lys Ala Gln Asn Leu Tyr Gln Glu Leu Leu Thr Gln Glu Gly
 145 4150 4155 4160
 Gln Ala Ser Phe Gln Gly Leu Lys Asp Asn Val Phe Asp Gly Leu Val
 4165 4170 4175
 Arg Val Thr Gln Glu Phe His Met Lys Val Lys His Leu Ile Asp Ser
 4180 4185 4190
 Leu Ile Asp Phe Leu Asn Phe Pro Arg Phe Gln Phe Pro Gly Lys Pro
 4195 4200 4205
 Gly Ile Tyr Thr Arg Glu Glu Leu Cys Thr Met Phe Ile Arg Glu Val
 4210 4215 4220
 Gly Thr Val Leu Ser Gln Val Tyr Ser Lys Val His Asn Gly Ser Glu
 225 4230 4235 4240
 Ile Leu Phe Ser Tyr Phe Gln Asp Leu Val Ile Thr Leu Pro Phe Glu
 4245 4250 4255
 Leu Arg Lys His Lys Leu Ile Asp Val Ile Ser Met Tyr Arg Glu Leu
 4260 4265 4270
 Leu Lys Asp Leu Ser Lys Glu Ala Gln Glu Val Phe Lys Ala Ile Gln
 4275 4280 4285
 Ser Leu Lys Thr Thr Glu Val Leu Arg Asn Leu Gln Asp Leu Leu Gln
 4290 4295 4300
 Phe Ile Phe Gln Leu Ile Glu Asp Asn Ile Lys Gln Leu Lys Glu Met
 305 4310 4315 4320
 Lys Phe Thr Tyr Leu Ile Asn Tyr Ile Gln Asp Glu Ile Asn Thr Ile
 4325 4330 4335
 Phe Asn Asp Tyr Ile Pro Tyr Val Phe Lys Leu Leu Lys Glu Asn Leu
 4340 4345 4350
 Cys Leu Asn Leu His Lys Phe Asn Glu Phe Ile Gln Asn Glu Leu Gln
 4355 4360 4365
 Glu Ala Ser Gln Glu Leu Gln Gln Ile His Gln Tyr Ile Met Ala Leu
 4370 4375 4380
 Arg Glu Glu Tyr Phe Asp Pro Ser Ile Val Gly Trp Thr Val Lys Tyr
 385 4390 4395 4400
 Tyr Glu Leu Glu Glu Lys Ile Val Ser Leu Ile Lys Asn Leu Leu Val
 4405 4410 4415
 Ala Leu Lys Asp Phe His Ser Glu Tyr Ile Val Ser Ala Ser Asn Phe
 4420 4425 4430
 Thr Ser Gln Leu Ser Ser Gln Val Glu Gln Phe Leu His Arg Asn Ile
 4435 4440 4445
 Gln Glu Tyr Leu Ser Ile Leu Thr Asp Pro Asp Gly Lys Gly Lys Glu

4450 4455 4460
 Lys Ile Ala Glu Leu Ser Ala Thr Ala Gln Glu Ile Ile Lys Ser Gln
 465 4470 4475 4480
 Ala Ile Ala Thr Lys Lys Ile Ile Ser Asp Tyr His Gln Gln Phe Arg
 4485 4490 4495
 Tyr Lys Leu Gln Asp Phe Ser Asp Gln Leu Ser Asp Tyr Tyr Glu Lys
 4500 4505 4510
 Phe Ile Ala Glu Ser Lys Arg Leu Ile Asp Leu Ser Ile Gln Asn Tyr
 4515 4520 4525
 His Thr Phe Leu Ile Tyr Ile Thr Glu Leu Leu Lys Lys Leu Gln Ser
 4530 4535 4540
 Thr Thr Val Met Asn Pro Tyr Met Lys Leu Ala Pro Gly Glu Leu Thr
 545 4550 4555 4560
 Ile Ile Leu

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14070 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GAGGAGCCCG CCCAGCCAGC CAGGGCCGCG AGGCCGAGGC CAGGCCGCGAG CCCAGGAGCC	60
GCCCCACCGC AGCTGGCGAT GGACCCGCCG AGGCCCGCGC TGCTGGCGCT GCTGGCGCTG	120
CCTGCGCTGC TGCTGCTGCT GCTGGCGGGC GCCAGGGCCG AAGAGGAAAT GCTGGAAAAT	180
GTCAGCCTGG TCTGTCCAAA AGATGCGACC CGATTCAAGC ACCTCCGGAA GTACACATAC	240
AACTATGAGG CTGAGAGTTC CAGTGGAGTC CCTGGGACTG CTGATTCAAG AAGTGCCACC	300
AGGATCAACT GCAAGGTTGA GCTGGAGGTT CCCAGCTCT GCAGCTTCAT CCTGAAGACC	360
AGCCAGTGCA TCCTGAAAGA GGTGTATGGC TTCAACCCTG AGGGCAAAGC CTTGCTGAAG	420
AAAACCAAGA ACTCTGAGGA GTTTGCTGCA GCCATGTCCA GGTATGAGCT CAAGCTGGCC	480
ATTCCAGAAG GGAAGCAGGT TTTCCTTTAC CCGGAGAAAG ATGAACCTAC TTACATCCTG	540
AACATCAAGA GGGGCATCAT TTCTGCCCTC CTGGTTCCCC CAGAGACAGA AGAAGCCAAG	600
CAAGTGTGT TTCTGGATAC CGTGTATGGA AACTGCTCCA CTCACTTTAC CGTCAAGACG	660
AGGAAGGGCA ATGTGGCAAC AGAAATATCC ACTGAAAGAG ACCTGGGGCA GTGTGATCGC	720
TTCAAGCCCA TCCGCACAGG CATCAGCCCA CTTGCTCTCA TCAAAGGCAT GACCCGCCCC	780
TTGTCAACTC TGATCAGCAG CAGCCAGTCC TGTCAGTACA CACTGGACGC TAAGAGGAAG	840
CATGTGGCAG AAGCCATCTG CAAGGAGCAA CACCTCTTCC TGCCTTTCTC CTACAAGAAT	900
AAGTATGGGA TGGTAGCACA AGTGACACAG ACTTTGAAAC TTGAAGACAC ACCAAAGATC	960
AACAGCCGCT TCTTTGGTGA AGGTACTAAG AAGATGGGCC TCGCATTTGA GAGCACCAA	1020
TCCACATCAC CTCCAAAGCA GGCCGAAGCT GTTTTGAAGA CTGTCCAGGA ACTGAAAAA	1080

CTAACCATCT	CTGAGCAAAA	TATCCAGAGA	GCTAATCTCT	TCAATAAGCT	GGTTACTGAG	1140
CTGAGAGGCC	TCAGTGATGA	AGCAGTCACA	TCTCTCTTGC	CACAGCTGAT	TGAGGTGTCC	1200
AGCCCCATCA	CTTTACAAGC	CTTGGTTCAG	TGTGGACAGC	CTCAGTGCTC	CACCTCACATC	1260
CTCCAGTGGC	TGAAACGTGT	GCATGCCAAC	CCCCTTCTGA	TAGATGTGGT	CACCTACCTG	1320
GTGGCCCTGA	TCCCCGAGCC	CTCAGCACAG	CAGCTGCGAG	AGATCTTCAA	CATGGCGAGG	1380
GATCAGCGCA	GCCGAGCCAC	CTTGTATGCG	CTGAGCCACG	CGGTCAACAA	CTATCATAAG	1440
ACAAACCCTA	CAGGGACCCA	GGAGCTGCTG	GACATTGCTA	ATTACCTGAT	GGAACAGATT	1500
CAAGATGACT	GCACTGGGGA	TGAAGATTAC	ACCTATTTGA	TTCTGCGGGT	CATTGGAAAT	1560
ATGGGCCAAA	CCATGGAGCA	GTAACTCCA	GAACTCAAGT	CTTCAATCCT	GAAATGTGTC	1620
CAAAGTACAA	AGCCATCACT	GATGATCCAG	AAAGCTGCCA	TCCAGGCTCT	GCGGAAAATG	1680
GAGCCTAAAG	ACAAGGACCA	GGAGGTTCTT	CTTCAGACTT	TCCTTGATGA	TGCTTCTCCG	1740
GGAGATAAGC	GACTGGCTGC	CTATCTTATG	TTGATGAGGA	GTCCTTCACA	GGCAGATATT	1800
AACAAAATTG	TCCAAATTCT	ACCATGGGAA	CAGAATGAGC	AAGTGAAGAA	CTTTGTGGCT	1860
TCCCATATTG	CCAATATCTT	GAAGCTCAGAA	GAATTGGATA	TCCAAGATCT	GAAAAAGTTA	1920
GTGAAAGAAG	TTCTGAAAGA	ATCTCAACTT	CCAACTGTCA	TGGACTTCAG	AAAATTCTCT	1980
CGGAATATC	AACTCTACAA	ATCTGTTTCT	ATTCCATCAC	TTGACCCAGC	CTCAGCCAAA	2040
ATAGAAGGGA	ATCTTATATT	TGATCCAAAT	AACTACCTTC	CTAAAGAAAG	CATGCTGAAA	2100
ACTACCCTCA	CTGCCTTTGG	ATTGCTTCA	GCTGACCTCA	TCGAGATTGG	CTTGGAAGGA	2160
AAAGGCTTTG	AGCCAACATT	GGAAGCTCTT	TTTGGGAAGC	AAGGATTTTT	CCCAGACAGT	2220
GTCAACAAAG	CTTTGTACTG	GGTTAATGGT	CAAGTTCCTG	ATGGTGTCTC	TAAGGTCTTA	2280
GTGGACCACT	TTGGCTATAC	CAAAGATGAT	AAACATGAGC	AGGATATGGT	AAATGGAATA	2340
ATGCTCAGTG	TTGAGAAGCT	GATTAAAGAT	TTGAAATCCA	AAGAAGTCCC	GGAAGCCAGA	2400
GCCTACCTCC	GCATCTTGGG	AGAGGAGCTT	GGTTTTGCCA	GTCTCCATGA	CCTCCAGCTC	2460
CTGGGAAAGC	TGCTTCTGAT	GGGTGCCCGC	ACTCTGCAGG	GGATCCCCCA	GATGATTGGA	2520
GAGGTCATCA	GGAAGGGCTC	AAAGAATGAC	TTTTTTCTTC	ACTACATCTT	CATGGAGAAT	2580
GCCTTTGAAC	TCCCCACTGG	AGCTGGATTA	CAGTTGCAAA	TATCTTCATC	TGGAGTCATT	2640
GCTCCCGGAG	CCAAGGCTGG	AGTAAACTG	GAAGTAGCCA	ACATGCAGGC	TGAACTGGTG	2700
GCAAAACCCT	CCGTGTCTGT	GGAGTTTGTG	ACAAATATGG	GCATCATCAT	TCCGGACTTC	2760
GCTAGGAGTG	GGGTCCAGAT	GAACACCAAC	TTCTTCCACG	AGTCGGGTCT	GGAGGCTCAT	2820
GTTGCCCTAA	AACCTGGGAA	GCTGAAGTTT	ATCATTCCTT	CCCCAAAGAG	ACCAGTCAAG	2880
CTGCTCAGTG	GAGGCAACAC	ATTACATTTG	GTCTCTACCA	CCAAAACGGA	GGTGATCCCA	2940
CCTCTCATTG	AGAACAGGCA	GTCCTGGTCA	GTTTGCAAGC	AAGTCTTTCC	TGGCCTGAAT	3000
TACTGCACCT	CAGGCGCTTA	CTCCAACGCC	AGCTCCACAG	ACTCCGCCTC	CTACTATCCG	3060
CTGACCGGGG	ACACCAGATT	AGAGCTGGAA	CTGAGGCCTA	CAGGAGAGAT	TGAGCAGTAT	3120
TCTGTCAGCG	CAACCTATGA	GCTCCAGAGA	GAGGACAGAG	CCTTGGTGGA	TACCCTGAAG	3180
TTTGTAATC	AAGCAGAAGG	TGCGAAGCAG	ACTGAGGCTA	CCATGACATT	CAAATATAAT	3240
CGGCAGAGTA	TGACCTTGTC	CAGTGAAGTC	CAAATTCCGG	ATTTTGATGT	TGACCTCGGA	3300
ACAATCCTCA	GAGTTAATGA	TGAATCTACT	GAGGGCAAAA	CGTCTTACAG	ACTCACCTTG	3360
GACATTCAGA	ACAAGAAAAT	TACTGAGGTC	GCCCTCATGG	GCCACCTAAG	TTGTGACACA	3420
AAGGAAGAAA	GAAAAATCAA	GGGTGTTATT	TCCATACCCC	GTTTGCAAGC	AGAAGCCAGA	3480
AGTGAGATCC	TCGCCCCTG	GTCGCCTGCC	AAACTGCTTC	TCCAAATGGA	CTCATCTGCT	3540
ACAGCTTATG	GCTCCACAGT	TTCCAAGAGG	GTGGCATGGC	ATTATGATGA	AGAGAAGATT	3600
GAATTTGAAT	GGAACACAGG	CACCAATGTA	GATACCAAAA	AAATGACTTC	CAATTTCCCT	3660
GTGGATCTCT	CCGATTATCC	TAAGAGCTTG	CATATGTATG	CTAATAGACT	CCTGGATCAC	3720
AGAGTCCCTC	AAACAGACAT	GACTTTCCGG	CACGTGGGTT	CCAAATTAAT	AGTTGCAATG	3780

AGCTCATGGC	TTCAGAAGGC	ATCTGGGAGT	CTTCCTTATA	CCCAGACTTT	GCAAGACCAC	3840
CTCAATAGCC	TGAAGGAGTT	CAACCTCCAG	AACATGGGAT	TGCCAGACTT	CCACATCCCA	3900
GAAAACCTCT	TCTTAAAAAG	CGATGGCCGG	GTCAAATATA	CCTTGAACAA	GAACAGTTTG	3960
AAAATTGAGA	TTCTTTTGCC	TTTTGGTGGC	AAATCCTCCA	GAGATCTAAA	GATGTTAGAG	4020
ACTGTTAGGA	CACCAGCCCT	CCACTTCAAG	TCTGTGGGAT	TCCATCTGCC	ATCTCGAGAG	4080
TTCCAAGTCC	CTACTTTTAC	CATTCCCAAG	TTGTATCAAC	TGCAAGTGCC	TCTCCTGGGT	4140
GTTCTAGACC	TCTCCACGAA	TGTCTACAGC	AACTTGATACA	ACTGGTCCGC	CTCCTACAGT	4200
GGTGGCAACA	CCAGCACAGA	CCATTTTCAGC	CTTCGGGCTC	GTTACCACAT	GAAGGCTGAC	4260
TCTGTGGTTG	ACCTGCTTTC	CTACAATGTG	CAAGGATCTG	GAGAAACAAC	ATATGACCAC	4320
AAGAATACGT	TCACACTATC	ATGTGATGGG	TCTCTACGCC	ACAAATTTCT	AGATTCGAAT	4380
ATCAAATTCA	GTCAATGTAG	AAAACCTGGA	AACAACCCAG	TCTCAAAGG	TTTACTAATA	4440
TTCGATGCAT	CTAGTTCCTG	GGGACCACAG	ATGTCTGCTT	CAGTTCATTT	GGACTCCAAA	4500
AAGAAACAGC	ATTTGTTTGT	CAAAGAAGTC	AAGATTGATG	GGCAGTTCAG	AGTCTCTTCG	4560
TTCTATGCTA	AAGGCACATA	TGGCCTGTCT	TGTCAGAGGG	ATCCTAACAC	TGGCCGGCTC	4620
AATGGAGAGT	CCAACCTGAG	GTTTAACTCC	TCCTACCTCC	AAGGCACCAA	CCAGATAACA	4680
GGAAGATATG	AAGATGGAAC	CCTCTCCCTC	ACCTCCACCT	CTGATCTGCA	AAGTGGCATC	4740
ATTAAAAATA	CTGCTTCCCT	AAAGTATGAG	AACTACGAGC	TGACTTTAAA	ATCTGACACC	4800
AATGGGAAGT	ATAAGAACTT	TGCCACTTCT	AACAAGATGG	ATATGACCTT	CTCTAAGCAA	4860
AATGCACTGC	TGCGTTCTGA	ATATCAGGCT	GATTACGAGT	CATTGAGGTT	CTTCAGCCTG	4920
CTTTCTGGAT	CACTAAATTC	CCATGGTCTT	GAGTTAAATG	CTGACATCTT	AGGCACTGAC	4980
AAAATTAATA	GTGGTGCTCA	CAAGGCGACA	CTAAGGATTG	GCCAAGATGG	AATATCTACC	5040
AGTGCAACGA	CCAACCTGAA	GTGTAGTCTC	CTGGTGCTGG	AGAATGAGCT	GAATGCAGAG	5100
CTTGGCCTCT	CTGGGGCATC	TATGAAATTA	ACAACAAATG	GCCGCTTCAG	GGAACACAAT	5160
GCAAAATTCA	GTCTGGATGG	GAAAGCCGCC	CTCACAGAGC	TATCACTGGG	AAGTGCTTAT	5220
CAGGCCATGA	TTCTGGGTGT	CGACAGCAAA	AACATTTTCA	ACTTCAAGGT	CAGTCAAGAA	5280
GGACTTAAGC	TCTCAAATGA	CATGATGGGC	TCATATGCTG	AAATGAAATT	TGACCACACA	5340
AACAGTCTGA	ACATTGCAGG	CTTATCACTG	GACTTCTCTT	CAAAACTTGA	CAACATTTAC	5400
AGCTCTGACA	AGTTTTATAA	GCAAACCTGTT	AATTTACAGC	TACAGCCCTA	TTCTCTGGTA	5460
ACTACTTTAA	ACAGTGACCT	GAAATACAAT	GCTCTGGATC	TCACCAACAA	TGGGAAACTA	5520
CGGCTAGAAC	CCCTGAAGCT	GCATGTGGCT	GGTAACCTAA	AAGGAGCCTA	CCAAAATAAT	5580
GAAATAAAAC	ACATCTATGC	CATCTCTTCT	GCTGCCTTAT	CAGCAAGCTA	TAAAGCAGAC	5640
ACTGTTGCTA	AGGTTTCAGG	TGTGGAGTTT	AGCCATCGGC	TCAACACAGA	CATCGCTGGG	5700
CTGGCTTCAG	CCATTGACAT	GAGCACAAAC	TATAATTTCAG	ACTCACTGCA	TTTCAGCAAT	5760
GTCTTCCGTT	CTGTAATGGC	CCCGTTTACC	ATGACCATCG	ATGCACATAC	AAATGGCAAT	5820
GGGAAACTCG	CTCTCTGGGG	AGAACATACT	GGGCAGCTGT	ATAGCAAATT	CCTGTTGAAA	5880
GCAGAACCTC	TGGCATTTAC	TTTCTCTCAT	GATTACAAAG	GCTCCACAAG	TCATCATCTC	5940
GTGTCTAGGA	AAAGCATCAG	TGCAGCTCTT	GAACACAAAG	TCAGTGCCCT	GCTTACTCCA	6000
GCTGAGCAGA	CAGGCACCTG	GAAACTCAAG	ACCCAATTTA	ACAACAATGA	ATACAGCCAG	6060
GACTTGATG	CTTACAACAC	TAAAGATAAA	ATTGGCGTGG	AGCTTACTGG	ACGAACTCTG	6120
GCTGACCTAA	CTCTACTAGA	CTCCCCAATT	AAAGTGCCAC	TTTTACTCAG	TGAGCCCATC	6180
AATATCATTG	ATGCTTTAGA	GATGAGAGAT	GCCGTTGAGA	AGCCCCAAGA	ATTTACAATT	6240
GTTGCTTTTG	TAAAGTATGA	TAAAAACCAA	GATGTTCACT	CCATTAACTT	CCCATTTTTT	6300
GAGACCTTGC	AAGAATATTT	TGAGAGGAAT	CGACAAACCA	TTATAGTTGT	ACTGGAAAAC	6360
GTACAGAGAA	ACCTGAAGCA	CATCAATATT	GATCAATTTG	TAAGAAAATA	CAGAGCAGCC	6420
CTGGGAAAAC	TCCCACAGCA	AGCTAATGAT	TATCTGAATT	CATTCAATTG	GGAGAGACAA	6480

GTTTCACATG	CCAAGGAGAA	ACTGACTGCT	CTCACAAAA	AGTATAGAAT	TACAGAAAAT	6540
GATATACAAA	TTGCATTAGA	TGATGCCAAA	ATCAACTTTA	ATGAAAAACT	ATCTCAACTG	6600
CAGACATATA	TGATACAATT	TGATCAGTAT	ATTAAAGATA	GTTATGATTT	ACATGATTTG	6660
AAAATAGCTA	TTGCTAATAT	TATTGATGAA	ATCATTGAAA	AATTAAAAAG	TCTTGATGAG	6720
CACTATCATA	TCCGTGTAAA	TTTAGTAAAA	ACAATCCATG	ATCTACATTT	GTTTATTGAA	6780
AATATTGATT	TTAACAAAAG	TGGAAGTAGT	ACTGCATCCT	GGATTCAAAA	TGTGGATACT	6840
AAGTACCAAA	TCAGAATCCA	GATACAAGAA	AACTGCAGC	AGCTTAAGAG	ACACATACAG	6900
AATATAGACA	TCCAGCACCT	AGCTGGAAAG	TTAAAACAAC	ACATTGAGGC	TATTGATGTT	6960
AGAGTGCTTT	TAGATCAATT	GGGAACTACA	ATTTCAATTTG	AAAGAATAAA	TGATGTTCTT	7020
GAGCATGTCA	AACACTTTGT	TATAAATCTT	ATTGGGGATT	TTGAAGTAGC	TGAGAAAATC	7080
AATGCCTTCA	GAGCCAAAGT	CCATGAGTTA	ATCGAGAGGT	ATGAAGTAGA	CCAACAAATC	7140
CAGGTTTTTA	TGGATAAATT	AGTAGAGTTG	GCCCACCAAT	ACAAGTTGAA	GGAGACTATT	7200
CAGAAGCTAA	GCAATGTCCT	ACAACAAGTT	AAGATAAAAG	ATTACTTTGA	GAAATTGGTT	7260
GGATTTATTG	ATGATGCTGT	CAAGAAGCTT	AATGAATTAT	CTTTTAAAC	ATTCAATTGAA	7320
GATGTTAACA	AATTCCTTGA	CATGTTGATA	AAGAAATTAA	AGTCATTTGA	TTACCACCAG	7380
TTTGTAGATG	AAACCAATGA	CAAAATCCGT	GAGGTGACTC	AGAGACTCAA	TGGTGAAATT	7440
CAGGCTCTGG	AACTACCACA	AAAAGCTGAA	GCATTAAAAAC	TGTTTTTAGA	GGAAACCAAG	7500
GCCACAGTTG	CAGTGTATCT	GGAAAGCCTA	CAGGACACCA	AAATAACCTT	AATCATCAAT	7560
TGGTTACAGG	AGGCTTTAAG	TTCAGCATCT	TTGGCTCACA	TGAAGGCCAA	ATTCCGAGAG	7620
ACTCTAGAAG	ATACACGAGA	CCGAATGTAT	CAAATGGACA	TTCAGCAGGA	ACTTCAACGA	7680
TACCTGTCTC	TGGTAGGCCA	GGTTTATAGC	ACACTTGTC	CCTACATTTT	TGATTGGTGG	7740
ACTCTTGCTG	CTAAGAACCT	TACTGACTTT	GCAGAGCAAT	ATTCTATCCA	AGATTGGGCT	7800
AAACGTATGA	AAGCATTGGT	AGAGCAAGGG	TTCACTGTTC	CTGAAATCAA	GACCATCCTT	7860
GGGACCATGC	CTGCCTTTGA	AGTCAGTCTT	CAGGCTCTTC	AGAAAGCTAC	CTTCCAGACA	7920
CCTGATTTTA	TAGTCCCCCT	AACAGATTTG	AGGATTCCAT	CAGTTCAGAT	AAACTTCAA	7980
GACTTAAAAA	ATATAAAAAAT	CCCATCCAGG	TTTTCCACAC	CAGAATTTAC	CATCCTTAAC	8040
ACCTTCCACA	TTCTTCTCTT	TACAATTGAC	TTTGTAGAAA	TGAAAGTAAA	GATCATCAGA	8100
ACCATTGACC	AGATGCTGAA	CAGTGAGCTG	CAGTGGCCCC	TTCCAGATAT	ATATCTCAGG	8160
GATCTGAAGG	TGGAGGACAT	TCCTCTAGCG	AGAATCACCC	TGCCAGACTT	CCGTTTACCA	8220
GAAATCGCAA	TTCCAGAATT	CATAATCCCA	ACTCTCAACC	TTAATGATTT	TCAAGTTCCT	8280
GACCTTCACA	TACCAGAATT	CCAGCTTCCC	CACATCTCAC	ACACAATTGA	AGTACCTACT	8340
TTTGGCAAGC	TATACAGTAT	TCTGAAAATC	CAATCTCCTC	TTTTCACATT	AGATGCAAAAT	8400
GCTGACATAG	GGAATGGAAC	CACCTCAGCA	AACGAAGCAG	GTATCGCAGC	TTCCATCACT	8460
GCCAAAGGAG	AGTCCAAATT	AGAAGTTCTC	AATTTTGATT	TTCAAGCAAA	TGCACAACCTC	8520
TCAAACCCTA	AGATTAATCC	GCTGGCTCTG	AAGGAGTCAG	TGAAGTTCTC	CAGCAAGTAC	8580
CTGAGAACGG	AGCATGGGAG	TGAAATGCTG	TTTTTTGGAA	ATGCTATTGA	GGGAAAATCA	8640
AACACAGTGG	CAAGTTTACA	CACAGAAAAA	AATACACTGG	AGCTTAGTAA	TGGAGTGATT	8700
GTCAAGATAA	ACAATCAGCT	TACCCTGGAT	AGCAACACTA	AATACTTCCA	CAAATTGAAC	8760
ATCCCCAAAC	TGGACTTCTC	TAGTCAGGCT	GACCTGCGCA	ACGAGATCAA	GACACTGTTG	8820
AAAGCTGGCC	ACATAGCATG	GACTTCTTCT	GGAAAAGGGT	CATGGAAATG	GGCCTGCCCC	8880
AGATTCTCAG	ATGAGGGAAC	ACATGAATCA	CAAATTAGTT	TCACCATAGA	AGGACCCCTC	8940
ACTTCTTTTG	GACTGTCCAA	TAAGATCAAT	AGCAAACACC	TAAGAGTAAA	CCAAAACCTG	9000
GTTTATGAAT	CTGGCTCCCT	CAACTTTTCT	AAACTTGAAA	TTCAATCACA	AGTCGATTCC	9060
CAGCATGTGG	GCCACAGTGT	TCTAACTGCT	AAAGGCATGG	CACTGTTTGG	AGAAGGGAAG	9120
GCAGAGTTTA	CTGGGAGGCA	TGATGCTCAT	TTAAATGGAA	AGGTTATTGG	AACTTTGAAA	9180

AATTCTCTTT TCTTTTCAGC CCAGCCATTT GAGATCACGG CATCCACAAA CAATGAAGGG	9240
AATTTGAAAG TTCGTTTTCC ATTAAGGTTA ACAGGGAAGA TAGACTTCCT GAATAACTAT	9300
GCACTGTTTC TGAGTCCCAG TGCCCAGCAA GCAAGTTGGC AAGTAAGTGC TAGGTTCAAT	9360
CAGTATAAGT ACAACCAAAA TTTCTCTGCT GGAAACAACG AGAACATTAT GGAGGCCCAT	9420
GTAGGAATAA ATGGAGAAGC AAATCTGGAT TTCTTAAACA TTCCTTTAAC AATTCCTGAA	9480
ATGCGTCTAC CTTACACAAT AATCACAACCT CCTCCACTGA AAGATTTCTC TCTATGGGAA	9540
AAAACAGGCT TGAAGGAATT CTTGAAAACG ACAAAGCAAT CATTTGATTT AAGTGTAATA	9600
GCTCAGTATA AGAAAAACAA ACACAGGCAT TCCATCACAA ATCCTTTGGC TGTGCTTTGT	9660
GAGTTTATCA GTCAGAGCAT CAAATCCTTT GACAGGCATT TTGAAAAAAA CAGAAACAAT	9720
GCATTAGATT TTGTCACCAA ATCCTATAAT GAAACAAAAA TTAAGTTTGA TAAGTACAAA	9780
GCTGAAAAAT CTCACGACGA GCTCCCCAGG ACCTTTCAAA TTCCTGGATA CACTGTTCCA	9840
GTTGTCAATG TTGAAGTGTC TCCATTACCC ATAGAGATGT CGGCATTCCG CTATGTGTTT	9900
CCAAAAGCAG TCAGCATGCC TAGTTTCTCC ATCCTAGGTT CTGACGTCCG TGTGCCTTCA	9960
TACACATTAA TCCTGCCATC ATTAGAGCTG CCAGTCCTTC ATGTCCCTAG AAATCTCAAG	10020
CTTTCTCTTC CAGATTTCAA GGAATTGTGT ACCATAAGCC ATATTTTAT TCCTGCCATG	10080
GGCAATATTA CCTATGATTT CTCCTTTAA TCAAGTGTCA TCACACTGAA TACCAATGCT	10140
GAACCTTTTA ACCAGTCAGA TATTGTTGCT CATCTCCTTT CTTCATCTTC ATCTGTCATT	10200
GATGCACTGC AGTACAAATT AGAGGGCACC ACAAGATTGA CAAGAAAAAG GGGATTGAAG	10260
TTAGCCACAG CTCTGTCTCT GAGCAACAAA TTTGTGGAGG GTAGTCATAA CAGTACTGTG	10320
AGCTTAACCA CGAAAAATAT GGAAGTGTCA GTGGCAACAA CCACAAAAGC CCAAATTCCA	10380
ATTTTGAGAA TGAATTTCAA GCAAGAACTT AATGGAAATA CCAAGTCAAA ACCTACTGTC	10440
TCTTCCTCCA TGGAATTTAA GTATGATTTT AATTCTTCAA TGCTGTACTC TACCGCTAAA	10500
GGAGCAGTTG ACCACAAGCT TAGCTTGGA AGCCTCACCT CTTACTTTTC CATTGAGTCA	10560
TCTACCAAAG GAGATGTCAA GGGTTCGGTT CTTTCTCGGG AATATTCAGG AACTATTGCT	10620
AGTGAGGCCA ACACTTACTT GAATTCCAAG AGCACACGGT CTTCAAGTGA GCTGCAGGGC	10680
ACTTCCAAAA TTGATGATAT CTGGAACCTT GAAGTAAAAG AAAATTTTGC TGGAGAAGCC	10740
ACACTCCAAC GCATATATTC CCTCTGGGAG CACAGTACGA AAAACCACTT ACAGCTAGAG	10800
GGCCTCTTTT TCACCAACGG AGAACATACA AGCAAAGCCA CCCTGGAACCT CTCTCCATGG	10860
CAAATGTCAG CTCTTGTTCA GGTCCATGCA AGTCAGCCCA GTTCCTTCCA TGATTTCCCT	10920
GACCTTGGCC AGGAAGTGGC CCTGAATGCT AACACTAAGA ACCAGAAGAT CAGATGGAAA	10980
AATGAAGTCC GGATTCATTC TGGGTCTTTC CAGAGCCAGG TCGAGCTTTC CAATGACCAA	11040
GAAAAGGCAC ACCTTGACAT TGCAGGATCC TTAGAAGGAC ACCTAAGGTT CCTCAAAAAT	11100
ATCATCCTAC CAGTCTATGA CAAGAGCTTA TGGGATTTCC TAAAGCTGGA TGTAAACCACC	11160
AGCATTGGTA GGAGACAGCA TCTTCGTGTT TCAACTGCCT TTGTGTACAC CAAAAACCCC	11220
AATGGCTATT CATTCTCCAT CCCTGTAAAA GTTTTGGCTG ATAAATTCAT TATTCCTGGG	11280
CTGAACTAA ATGATCTAAA TTCAGTTCTT GTCATGCCTA CGTTCCATGT CCCATTTACA	11340
GATCTTCAGG TTCCATCGTG CAACTTGAC TTCAGAGAAA TACAAATCTA TAAGAAGCTG	11400
AGAACTTCAT CATTTGCCCT CAACCTACCA ACACTCCCCG AGGTAAAATT CCCTGAAGTT	11460
GATGTGTTAA CAAAATATTC TCAACCAGAA GACTCCTTGA TTCCCTTTTT TGAGATAACC	11520
GTGCCTGAAT CTCAGTTAAC TGTGTCCCAG TTCACGCTTC CAAAAAGTGT TTCAGATGGC	11580
ATTGCTGCTT TGGATCTAAA TGCAAGTAGCC AACAAGATCG CAGACTTTGA GTTGCCACCC	11640
ATCATCGTGC CTGAGCAGAC CATTGAGATT CCCTCCATTA AGTTCTCTGT ACCTGCTGGA	11700
ATTGCCATTC CTTCTTTTCA AGCACTGACT GCACGCTTTG AGGTAGACTC TCCCGTGTAT	11760
AATGCCACTT GGAGTGCCAG TTTGAAAAAC AAAGCAGATT ATGTTGAAAC AGTCCTGGAT	11820
TCCACATGCA GCTCAACCGT ACAGTTCCTA GAATATGAAC TTAATGTTTT GGGAAACACAC	11880

AAAATCGAAG ATGGTACGTT AGCCTCTAAG ACTAAAGGAA CATTTCACACA CCGTGA	11940
AGTGCAGAAT ATGAAGAAGA TGGCAAATAT GAAGGACTTC AGGAATGGGA AGGAAAAGCG	12000
CACCTCAATA TCAAAAGCCC AGCGTTCACC GATCTCCATC TGCGCTACCA GAAAGACAAG	12060
AAAGGCATCT CCACCTCAGC AGCCTCCCCA GCCGTAGGCA CCGTGGGCAT GGATATGGAT	12120
GAAGATGACG ACTTTTCTAA ATGGAACCTC TACTACAGCC CTCAGTCCTC TCCAGATAAA	12180
AAACTCACCA TATTCAAAAC TGAGTTGAGG GTCCGGGAAT CTGATGAGGA AACTCAGATC	12240
AAAGTTAATT GGAAGAAGA GGCAGCTTCT GGCTTGCTAA CCTCTCTGAA AGACAACGTG	12300
CCCAAGGCCA CAGGGGTCCT TTATGATTAT GTCAACAAGT ACCACTGGGA ACACACAGGG	12360
CTCACCTGA GAGAAGTGTC TTCAAAGCTG AGAAGAAATC TGCAGGACCA TGCTGAGTGG	12420
GTTTATCAAG GGGCCATTAG GGAAATTGAT GATATCGACG AGAGGTTCCA GAAAGGAGCC	12480
AGTGGGACCA CTGGGACCTA CCAAGAGTGG AAGGACAAGG CCCAGAATCT GTACCAGGAA	12540
CTGTTGACTC AGGAAGGCCA AGCCAGTTTC CAGGGACTCA AGGATAACGT GTTTGATGGC	12600
TTGGTACGAG TTA	12660
GATTTTCTGA ACTTCCCCAG ATTCCAGTTT CCGGGGAAAC CTGGGATATA CACTAGGGAG	12720
GAACTTTGCA CTATGTTTCA	12780
GTCCATAATG GTTCAGAAAT ACTGTTTTCC TATTTCCAAG ACCTAGTGAT TACACTTCCT	12840
TTCGAGTTAA GGAAACATAA ACTAATAGAT GTAATCTCGA TGTATAGGGA ACTGTTGAAA	12900
GATTTATCAA AAGAAGCCCC AGAGGTATTT AAAGCCATTC AGTCTCTCAA GACCACAGAG	12960
GTGCTACGTA ATCTTCAGGA CCTTTTACAA TTCATTTTCC AACTAATAGA AGATAACATT	13020
AAACAGCTGA AAGAGATGAA ATTTACTTAT CTTATTAATT ATATCCAAGA TGAGATCAAC	13080
ACAATCTTCA ATGATTATAT CCCATATGTT TTTAAATTGT TGAAAGAAAA CCTATGCCTT	13140
AATCTTCATA AGTTCAATGA ATTTATTCAA AACGAGCTTC AGGAAGCTTC TCAAGAGTTA	13200
CAGCAGATCC ATCAATACAT TATGGCCCTT CGTGAAGAAT ATTTTGATCC AAGTATAGTT	13260
GGCTGGACAG TGAAATATTA TGA	13320
TTAGTTGCTC TTAAGGACTT CCATTCTGAA TATATTGTCA GTGCCTCTAA CTTTACTTCC	13380
CAACTCTCAA GTCAAGTTGA GCAATTTCTG CACAGAAATA TTCAGGAATA TCTTAGCATC	13440
CTTACCGATC CAGATGGAAA AGGGAAAGAG AAGATTGCAG AGCTTTCTGC CACTGCTCAG	13500
GAAATAATTA AAAGCCAGGC CATTGCGACG AAGAAAATAA TTTCTGATTA CCACCAGCAG	13560
TTAGATATA AACTGCAAGA TTTTTCAGAC CAACTCTCTG ATTACTATGA AAAATTTATT	13620
GCTGAATCCA AAAGATTGAT TGACCTGTCC ATTCAAACT ACCACACATT TCTGATATAC	13680
ATCACGGAGT TACTGAAAA GCTGCAATCA ACCACAGTCA TGAACCCCTA CATGAAGCTT	13740
GCTCCAGGAG AACTTACTAT CATCCTCTAA TTTTAAAA GAAATCTTCA TTTATTCTTC	13800
TTTTCCAATT GAACTTTAC ATAGCACAGA AAAAATTCAA AATGCCTATA TTGATCAAAC	13860
CATACAGTGA GCCAGCCTTG CAGTAGGCAG TAGACTATAA GCAGAAGCAC ATATGAACTG	13920
GACCTGCACC AAAGCTGGCA CCAGGGCTCG GAAGGTCTCT GAACTCAGAA GGATGGCATT	13980
TTTTGCAAGT TAAAGAAAAT CAGGATCTGA GTTATTTTGC TAAACTGGG GGAGGAGGAA	14040
CAAATAAATG GAGTCTTTAT TGTGTATCAT	14070

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3805 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(ix) FEATURE:

(A) NAME/KEY: exon

(B) LOCATION: 71...114

(D) OTHER INFORMATION: Exon 1

(A) NAME/KEY: exon

(B) LOCATION: 872...937

(D) OTHER INFORMATION: Exon 2

(A) NAME/KEY: exon

(B) LOCATION: 2031...2223

(D) OTHER INFORMATION: Exon 3

(A) NAME/KEY: exon

(B) LOCATION: 2805...3664

(C) OTHER INFORMATION: Exon 4

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CCTATCCCTG	GGGGAGGGGG	CGGGACAGGG	GGAGCCCTAT	AATTGGACAA	GTCTGGGATC	60
CTTGAGTCCT	ACTCAGCCCC	AGCGGAGGTG	AAGGACGTCC	TTCCCCAGGA	GCCGGTGAGA	120
AGCGCAGTCG	GGGGCACGGG	GATGAGCTCA	GGGGCCTCTA	GAAAGAGCTG	GGACCCTGGG	180
AAGCCCTGGC	CTCCAGGTAG	TCTCAGGAGA	GCTACTCGGG	GTCGGGCTTG	GGGAGAGGAG	240
GAGCGGGGGT	GAGGCAAGCA	GCAGGGGACT	GGACCTGGGA	AGGGCTGGGC	AGCAGAGACG	300
ACCCGACCCG	CTAGAAGGTG	GGGTGGGGAG	AGCAGCTGGA	CTGGGATGTA	AGCCATAGCA	360
GGACTCCACG	AGTTGTCACT	ATCATTATCG	AGCACCTACT	GGGTGTCCCC	AGTGTCTCTA	420
GATCTCCATA	ACTGGGGAGC	CAGGGGCAGC	GACACGGTAG	CTAGCCGTCG	ATTGGAGAAC	480
TTTAAATGA	GGACTGAATT	AGCTCATAAA	TGGAACACGG	CGCTTAACTG	TGAGGTTGGA	540
GCTTAGAATG	TGAAGGGAGA	ATGAGGAATG	CGAGACTGGG	ACTGAGATGG	AACCGGCGGT	600
GGGGAGGGGG	TGGGGGGATG	GAATTTGAAC	CCCGGGAGAG	GAAGATGGAA	TTTTCTATGG	660
AGGCCGACCT	GGGGATGGGG	AGATAAGAGA	AGACCAGGAG	GGAGTTAAAT	AGGGAATGGG	720
TTGGGGGCGG	CTTGGTAAAT	GTGCTGGGAT	TAGGCTGTTG	CAGATAATGC	AACAAGGCTT	780
GGAAGGCTAA	CCTGGGGTGA	GGCCGGGTTG	GGGGCGCTGG	GGGTGGGAGG	AGTCCTCACT	840
GGCGGTTGAT	TGACAGTTTC	TCCTTCCCCA	GA CTGGCCAA	TCACAGGCAG	GAAGATGAAG	900
GTTCTGTGGG	CTGCGTTGCT	GGTCACATTC	CTGGCAGGTA	TGGGGGCGGG	GCTTGCTCGG	960
TTCCCCCGCG	TCCTCCCCCT	CTCATCCTCA	CCTCAACCTC	CTGGCCCCAT	TCAGACAGAC	1020
CCTGGGCCCC	CTCTTCTGAG	GCTTCTGTGC	TGCTTCCTGG	CTCTGAACAG	CGATTTGACG	1080
CTCTCTGGGC	CTCGGTTTCC	CCCATCCTTG	AGATAGGAGT	TAGAAGTTGT	TTGTTGTTG	1140
TTGTTTGTG	TTGTTGTTTT	GTTTTTTTGA	GATGAAGTCT	CGCTCTGTCT	CCCAGGCTGG	1200
AGTGCACTGG	CGGGATCTCG	GCTCACTGCA	AGCTCCGCCT	CCCAGGTCCA	CGCCATTCTC	1260

CTGCCTCAGC	CTCCCAAGTA	GCTGGGACTA	CAGGCACATG	CCACCACACC	CGACTAACTT	1320
TTTTGTATTT	TCAGTAGAGA	CGGGGTTTCA	CCATGTTGGC	CAGGCTGGTC	TGGAATCCT	1380
GACCTCAGGT	GATCTGCCCC	TTTCGATCTC	CCAAAGTGCT	GGGATTACAG	GCGTGAGCCA	1440
CCGCACCTGG	CTGGGAGTTA	GAGGTTTCTA	ATGCATTGCA	GGCAGATAGT	GAATACCAGA	1500
CACGGGGCAG	CTGTGATCTT	TATTCTCCAT	CACCCCCACA	CAGCCCTGCC	TGGGGCACAC	1560
AAGGACACTC	AATACATGCT	TTTCCGCTGG	GCCGGTGGCT	CACCCCTGTA	ATCCCAGCAC	1620
TTTGGGAGGC	CAAGGTGGGA	GGATCACTTG	AGCCCAGGAG	TTCAACACCA	GCCTGGGCAA	1680
CATAGTGAGA	CCCTGTCTCT	ACTAAAAATA	CAAAAATTAG	CCAGGCATGG	TGCCACACAC	1740
CTGTGCTCTC	AGCTACTCAG	GAGGCTGAGG	CAGGAGGATC	GCTTGAGCCC	AGAAGGTCAA	1800
GGTTGCAGTG	AACCATGTTT	AGGCCGCTGC	ACTCCAGCCT	GGGTGACAGA	GCAAGACCCT	1860
GTTTATAAAT	ACATAATGCT	TTCCAAGTGA	TTAAACCGAC	TCCCCCTCA	CCCTGCCCCAC	1920
CATGGCTCCA	AAGAAGCATT	TGTGGAGCAC	CTTCTGTGTG	CCCCTAGGTA	GCTAGATGCC	1980
TGGACGGGGT	CAGAAGGACC	CTGACCCGAC	CTTGAACCTG	TTCCACACAG	GATGCCAGGC	2040
CAAGGTGGAG	CAAGCGGTGG	AGACAGAGCC	GGAGCCCGAG	CTGCGCCAGC	AGACCGAGTG	2100
GCAGAGCGGC	CAGCGCTGGG	AACTGGCACT	GGGTCGCTTT	TGGGATTACC	TGCGCTGGGT	2160
GCAGACACTG	TCTGAGCAGG	TGCAGGAGGA	GCTGCTCAGC	TCCCAGGTCA	CCCAGGAAct	2220
GAGGTGAGTG	TCCCCATCCT	GGCCCTTGAC	CCTCCTGGTG	GGCGGCTATA	CCTCCCCAGG	2280
TCCAGGTTTC	ATTCTGCCCC	TGTCGCTAAG	TCTTGGGGGG	CCTGGGTCTC	TGCTGGTTCT	2340
AGCTTCCTCT	TCCCATTCT	GACTCCTGGC	TTTAGCTCTC	TGGAATTCTC	TCTCTCAGCT	2400
TTGTCTCTCT	CTCTTCCCTT	CTGACTCAGT	CTCTCACACT	CGTCTGGCT	CTGTCTCTGT	2460
CCTTCCCTAG	CTCTTTTATA	TAGAGACAGA	GAGATGGGGT	CTCACTGTGT	TGCCCAGGCT	2520
GGTCTTGAAC	TTCTGGGCTC	AAGCGATCCT	CCCGCCTCGG	CCTCCCAAAG	TGCTGGGATT	2580
AGAGGCATGA	GCACCTTGCC	CGGCCTCCTA	GCTCCTTCTT	CGTCTCTGCC	TCTGCCCTCT	2640
GCATCTGCTC	TCTGCATCTG	TCTCTGTCTC	CTTCTCTCGG	CCTCTGCCCC	GTTCCTTCTC	2700
TCCCTCTTGG	GTCTCTCTGG	CTCATCCCCA	TCTCGCCCGC	CCCATCCCAG	CCCTTCTCCC	2760
CCGCCTCCCC	ACTGTGCGAC	ACCCTCCCGC	CCTCTCGGCC	GCAGGGCGCT	GATGGACGAG	2820
ACCATGAAGG	AGTTGAAGGC	CTACAAATCG	GAACGGAGG	AACAACCTGAC	CCCGTGGCGG	2880
GAGGAGACGC	GGGCACGGCT	GTCCAAGGAG	CTGCAGGCGG	CGCAGGCCCG	GCTGGGCGCG	2940
GACATGGAGG	ACGTGTGCGG	CCGCCTGGTG	CAGTACCGCG	GCGAGGTGCA	GGCCATGCTC	3000
GGCCAGAGCA	CCGAGGAGCT	GCGGGTGCGC	CTCGCCTCCC	ACCTGCGCAA	GCTGCGTAAG	3060
CGGCTCCTCC	GCGATGCCGA	TGACCTGCAG	AAGCGCCTGG	CAGTGTACCA	GGCCGGGGCC	3120
CGCGAGGGCG	CCGAGCGCGG	CCTCAGCGCC	ATCCGCGAGC	GCCTGGGGCC	CCTGGTGGAA	3180
CAGGGCCGCG	TGCGGGCCGC	CACTGTGGGC	TCCCTGGCCG	GCCAGCCGCT	ACAGGAGCGG	3240
GCCCAGGCCCT	GGGGCGAGCG	GCTGCGCGCG	CGGATGGAGG	AGATGGGCAG	CCGGACCCGC	3300
GACCGCCTGG	ACGAGGTGAA	GGAGCAGGTG	GCGGAGGTGC	GCGCCAAGCT	GGAGGAGCAG	3360
GCCCAGCAGA	TACGCCTGCA	GGCCGAGGCC	TTCAGGCC	GCCTCAAGAG	CTGGTTCGAG	3420
CCCCTGGTGG	AAGACATGCA	GCGCCAGTGG	GCCGGGCTGG	TGGAGAAGGT	GCAGGCTGCC	3480
GTGGGCACCA	GCGCCGCCCC	TGTGCCCAGC	GACAATCACT	GAACGCCGAA	GCCTGCAGCC	3540
ATGCGACCCC	ACGCCACCCC	GTGCCTCCTG	CCTCCGCGCA	GCCTGCAGCG	GGAGACCCTG	3600
TCCCCGCCCC	AGCCGTCCCTC	CTGGGGTGGG	CCCTAGTTTA	ATAAAGATT	ACCAAGTTTC	3660
ACGCATCTGC	TGGCCTCCCC	CTGTGATTTT	CTCTAAGCCC	CAGCCTCAGT	TTCTCTTTCT	3720
GCCCACATAC	TGCCACACAA	TTCTCAGCCC	CCTCCTCTCC	ATCTGTGTCT	GTGTGTATCT	3780
TTCTCTCTGC	CCTTTTTTTT	TTTTT				3805

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AGTCTGGATG GGTAAGCCGC CCTCA

25

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CTGGGCTGGC TTAAGCCATT GACAT

25

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GCTCTCTGGG GATAACATAC TGGGC

25

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GATGCCGTTG AGTAGCCCCA AGAAT

25

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GAGAGGAATC GATAAACCAT TATAG

25

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TGTAAGAAAA TAAAGAGCAG CCCTG

25

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GCAGCCCTGG GATAACTCCC ACAGC

25

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GCAAGCTAAT GATTAGCTGA ATTCATTCAA T

31

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CAAGTTTCAC ATGCCTAGGA GAAACTGACT G

31

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

ATATACAAAT TGCATGAGAT GATGCCAAAA T

31

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AAACTATCTC AACTGTAGAC ATATATGATA C

31

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GCTAATATTA TTGATTAAAT CATTGAAATT A

31

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

TGATGAGCAC TAGCATATCC GTGTA

25

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CTGCAGCAGC TTTAGAGACA CATAC

25

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

AACAGTGAGC TGTAGTGGCC CGTTC

25

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CAGACTTCCG TTAACCAGAA ATCGC

25

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

AAAGGGTCAT GGTAATGGGC CTGCC

25

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

ACATATATGA TATAATTTGA TCA GT

25

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

ATGGAGGACG TGTGCGGCCG CCTGG

25

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GACCTGCAGA AGTGCCTGGC AGTGT

25

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GACCTGCAGA AGCGCCTGGC AGTGT

25

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TAAGGTCAGG AGTTTGAGAC CAGCC

25

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

CTCGGAGAGC CCCCTCGCA

19

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CAAGGAGATA GTGGGGGAC

19

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

ACCATCGACG AGAAAGGGA

19

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

TTTGGACAGC GTCCATACT

19

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

TGCCTCGCCC AGGTCCTGG

19

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

CCCACTGCCA GGTATGGGC

19

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

AAAGATTCAT GTGAAGGAGA TAGTGGGGGA CCCCATGTT

39

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 13 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Lys Asp Ser Cys Glu Gly Asp Ser Gly Gly Pro His Val
1 5 10

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 39 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

AAAGATTCAT GTGAAGGAGA TCGTGGGGGA CCCCATGTT

39

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 68 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GGGGTCCCCC ACGATCTCCT TCACATTTTU GUGAAGGAGA TCGTGGGGGA CCCC GCGCGT 60
TTTCGCGC 68

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 68 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

TGTGAAGGAG ATCGTGGGGG ACCCCTTTTG GGGUCCCCCA CGATCUCCUU CACAGCGCGT 60
TTTCGCGC 68

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 68 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

TGTGAAGGAG ATCGTGGGGG ACCCCTTTTG GGGUCCCCCA CGATCUCCUU CACAGCGCGT 60
TTTCGCGC 68

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 68 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

TGTCAAGGAG ATCGTGGGGG ACCCCTTTTG GGGUCCCCCA CGATCUCCUU GACAGCGCGT 60
TTTCGCGC 68

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

CATTGCTGAC AAGGAATACA CGAAC 25

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

ATTTGCCTTT CATTGCACAC TCTTC 25

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

ATTGCCTTGC TGGAAGTGA TAAC

24

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

TTGCCTTTCA TTGCACATTC TTCAC

25

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

AAGGAGATAG TGGGGGA

17

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

AAGGAGATCG TGGGGGA

17

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

AAGGAGATAG TGGGGGA

17

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

AAGGAGATCG TGGGGGA

17

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

ATTGCCTTGC TGGAAGTGA TAAAC

25

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

TTGCCTTTCA TTGCACATTC TTCAC

25

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

AAGGAGATAG TGGGGGA

17

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

AAGGAGATCG TGGGGGA

17

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

GTTGACCGAG CCACATGCCT TAG

23

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 68 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

CTGGGCTGGC TTAAGCCATT GACATUUUUA UGUCAAUGGC UUAAGCCAGC CCAGGCGCGU
UUUCGCGC

60

68

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 88 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

GACAAGTTTC ACATGCCTAG GAGAACTGA CTGCTUUUUA GCAGUCAGUU UCUCCTAGGC
AUGUGAAACU UGUCGCGCGU UUUCGCGC

60

88

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 76 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

CCCTGAATGT CCTGGAAATG ACTGCCGATT TTTAUCTTCA GUCATTTCCA GGACAUUCAG 60
GGGCGCGTTT TCGCGC 76

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 80 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

TAAGTTTCTT AACTGGGATT ATTAGCTGGG GTTTTCCCCA GCUAATAATC CCAGTTAAGA 60
AACUUAGCGC GTTTTCGCGC 80

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 68 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

GTTTCTTAAC TGGGATTATT ACGTGT³⁵TTTTC AGCUAA⁴UAAT CCCAGUUAAG AAACGCGCGT 60
TTTCGCGC 68

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 80 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

TAAGTTTCTT AACTGGGATT ATTAGCTGGG GUUUUCCCCA GCUAAUAAUC CCAGUUAAGA 60
AACUUAGCGC GUUUUCGCGC 80

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 80 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

TAAGTTTCTT AACTGGGATT ATTAGCTGGG GUUUUCCCCA GCUAAUAAUC UCAGUUAAGA 60
AACUUAGCGC GUUUUCGCGC 80

CLAIMS:

1. A composition comprising:
 - a) a recombinagenic oligonucleobase;
 - b) an aqueous carrier; and
 - c) a macromolecular carrier selected from the group consisting of
 - (i) an aqueous-cored lipid vesicle, wherein the aqueous core contains the oligonucleobase,
 - (ii) a lipid nanosphere, which comprises a lipophilic salt of the oligonucleobase, and
 - (iii) a polycation having an average molecular weight of between 500 daltons and 1.3 Md wherein the polycation forms a salt with the oligonucleobase.
2. The composition of claim 1, in which the macromolecular carrier further comprises a ligand for a clathrin-coated pit receptor.
3. The composition of claim 2, in which the macromolecular carrier is a negatively charged, aqueous-cored lipid vesicle.
4. The composition of claim 3, in which the aqueous-cored lipid vesicle comprises dioleoylphosphatidylcholine and dioleoylphosphatidylserine.
5. The composition of claim 4, in which the aqueous-cored lipid vesicle further comprises a cerebroside.
6. The composition of claim 2, wherein the macromolecular carrier is a branched-chain polyethylenimine.
7. The composition of claim 2, wherein the macromolecular carrier is a linear polyethylenimine.
8. The composition of claim 2, wherein the macromolecular carrier is an aqueous-cored lipid vesicle that comprises a fusigenic F-protein.
9. The composition of claim 2 in which the oligonucleobase comprises:
 - (i) a first and a second homologous region that are together at least 16 and not more than 60 nucleobases in length, which regions are homologous with a target gene of a mammal; and
 - (ii) a heterologous region that is disposed between the first and second homologous region and is at least 1 and not more than 20 nucleobases in length, which is heterologous with the target gene and which contains the alteration.

10. The composition of claim 9, in which the recombinagenic oligonucleobase comprises at least 15 deoxynucleotides that are Watson-Crick base paired with 2'-Substituted Ribonucleotides.
11. The composition of claim 10, in which the 2'-Substituted Ribonucleotides are independently selected from the group consisting of 2'-methoxy-ribonucleotides, 2'-allyloxy-ribonucleotides, 2'-methoxyethoxy-ribonucleotides and 2'-fluoro-ribonucleotides.
12. The composition of claim 9, in which the ligand is a ligand for a receptor selected from the group consisting of the receptors for transferrin, nicotinic acid, carnitine, insulin and insulin like growth factor-1.
13. The composition of claim 9, in which the clathrin-coated pit receptor is an asialoglycoprotein receptor.
14. The composition of claim 13, in which the ligand comprises a moiety selected from the group consisting of lactose, galactose, and N-acetylgalactosamine, and in which the sequence of the oligonucleobase comprises the sequence of a contiguous 16 nucleotide fragment of a human gene that encodes a product selected from the group consisting of α 1-antitrypsin, coagulation factor IX, uridinediphosphoglucuronate glucuronosyltransferase, glucocerebrosidase, glucose-6-phosphatase, low density lipoprotein receptor, ornithine transcarbamylase and phenylalanine hydroxylase or the complement thereof.
15. A method of altering a target gene in a tissue of a subject mammal comprising administering to the subject mammal a composition of claim 1.
16. The method of claim 15, wherein the tissue is the liver.
17. The method of claim 15, wherein the oligonucleobase comprises:
 - a) a first and a second homologous region that are together at least 16 and not more than 60 nucleobases in length, which regions are homologous with the target gene of a mammal; and
 - b) a heterologous region that is disposed between the first and second homologous region and is at least 1 and not more than 20 nucleobases in length, which is heterologous with the target gene and which contains the alteration.
18. A method of treating a disease caused by a mutated sequence in a target gene in a cell in a human subject which comprises altering the mutated sequence in a number of cells in the subject effective to ameliorate the disease.

19. The method of claim 18, which further comprises the steps of determining the phenotypic effect of the altered target genes in the subject and subsequently increasing or decreasing said phenotypic effect by adjusting the number of said altered target genes in the subject.
20. The method of claim 18, wherein the cell is a hepatocyte.
21. The method of claim 18, which comprises administering to the subject an amount of a composition comprising:
- a) a recombinagenic oligonucleobase
 - b) an aqueous carrier; and
 - c) a macromolecular carrier selected from the group consisting of
 - (i) an aqueous-cored lipid vesicle, wherein the aqueous core contains the oligonucleobase,
 - (ii) a lipid nanosphere, which comprises a lipophilic salt of the oligonucleobase, and
 - (iii) a polycation having an average molecular weight of between 500 daltons and 1.3 Md wherein the polycation forms a salt with the oligonucleobase,
- wherein the macromolecular carrier further comprises a ligand for a clathrin-coated pit receptor and wherein the amount is effective to ameliorate a disease caused by the sequence.
22. The method of claim 21, wherein the oligonucleobase comprises:
- a) a first and a second homologous region that are together at least 16 and not more than 60 nucleobases in length, which regions are homologous with a target gene of a mammal; and
 - b) a heterologous region that is disposed between the first and second homologous region and is at least 1 and not more than 20 nucleobases in length, which is heterologous with the target gene and which contains the alteration.
23. The method of claim 18, wherein the target gene is an allele of the human gene that encodes a product selected from the group consisting of α 1-antitrypsin, coagulation factor IX, uridinediphosphoglucuronate glucuronosyltransferase, glucocerebrosidase, glucose-6-phosphatase, low density lipoprotein receptor, ornithine transcarbamylase and phenylalanine hydroxylase or the complement thereof.

24. A method of reducing LDL in the blood of a subject comprising altering an Apo B gene of a hepatocyte of the subject such that the transcript of the altered Apo B gene contains an in-frame stop codon whereby the altered gene encodes a protein having at least 1433 amino acids and not more than 3974 amino acids.
25. The method of claim 24, which further comprises the steps of determining effect on the level of LDL of the alteration of the Apo B genes in the subject and subsequently adjusting the number of altered Apo B genes in the subject.
26. The method of claim 24, wherein the altered gene encodes a protein having at least 1841 amino acids and not more than 2975 amino acids.
27. The method of claim 26, wherein the altered gene encodes a protein having a sequence of a fragment of SEQ ID No. 1, which fragment is at least amino acids 1-1841 and not more than amino acids 1-2975.
28. The method of claim 24, which comprises administering a recombinagenic oligonucleobase which comprises a first and a second homologous region each having a sequence of at least 10 nucleobases selected from nt 4342-11913 of SEQ ID No: 2, whereby the alteration of the Apo B gene is effected.
29. The method of claim 24, wherein the subject's fasting LDL serum cholesterol is reduced to below 140 mg/dl.
30. A composition for the modification of a human Apo B gene comprising an oligonucleobase which oligonucleobase comprises:
 - a) a first and a second homologous region that are each at least 8 nucleobases in length and together at least 20 nucleobases in length, which homologous regions are each homologous with a fragment of the sequence of nt 5649-9051 of SEQ ID No. 2, and
 - b) a heterologous region that is disposed between the first and second homologous region,such that the introduction of the sequence of the heterologous region into the Apo B gene results in the truncation of the protein encoded thereby.
31. The composition of claim 30, in which the first and the second homologous regions each comprises at least 3 contiguous nucleobase-pairs of hybrid-duplex.
32. The composition of claim 30, in which the sum of the lengths of the first and second homologous regions is not more than 60 nucleobases in length.
33. The composition of claim 30, in which the homologous regions together comprise between 9 and 25 nucleobase pairs of hybrid-duplex.

34. The composition of claim 30, in which the GC fraction of each homologous region is at least 33%.
35. The composition of claim 30, in which the GC fraction of a homologous region is at least 50%.
36. The composition of claim 30, in which the sequence of the oligonucleobase comprises the sequence of at least a 21 nucleobase fragment of any one of SEQ ID No. 4-20 or the complement thereof.
37. The composition of claim 30, in which the sequence of the oligonucleobase comprises the sequence of at least a 25 nucleobase fragment of any one of SEQ ID No. 4-20 or the complement thereof.
38. A method of treatment and/or prophylaxis in a subject comprising altering an Apo E gene of a hepatocyte of the subject by introducing a substitution selected from the group (Arg→Cys)¹¹², (Arg→Cys)¹⁵⁸ and (Cys→Arg)¹⁵⁸.
39. The method of claim 38, wherein the subject is homozygous for Apo E4 and the alteration comprises the substitution (Arg→Cys)¹¹².
40. The method of claim 39, which comprises administering a chimeric mutational vector having a sequence which comprises SEQ ID No: 22.
41. The method of claim 38, wherein the treatment or prophylaxis comprises reducing the subject's fasting serum LDL cholesterol level and the alteration comprises the substitution (Arg→Cys)¹⁵⁸.
42. The method of claim 41, which comprises administering a chimeric mutational vector having a sequence which comprises SEQ ID No: 23.
43. The method of claim 38, wherein the subject suffers from Type III hyperlipidemia and the alteration comprises the substitution (Cys→Arg)¹⁵⁸.
44. The method of claim 43, which comprises administering a recombinagenic oligonucleobase having a sequence which comprises SEQ ID No: 24.
45. A composition for the alteration of a human Apo E gene comprising a recombinagenic oligonucleobase having a sequence comprising the sequence of at least a 21 nucleobase fragment of any one of SEQ ID No. 22 - 25 or the complement thereof.

ABSTRACT

The present invention concerns compositions and methods for the introduction of specific genetic changes in endogenous genes of the cells of an animal. The genetic changes are effected by oligonucleotides or oligonucleotide derivatives and analogs, which are generally less than about 100 nucleotides in length. The invention provides for macromolecular carriers, optionally incorporating ligands for clathrin coated pit receptors. In one embodiment the ligand is a lactose or galactose and the genetic changes are made in hepatocytes. By means of the invention up to 40% of the copies of a target gene have been changed *in vitro*. Repair of mutant genes having a Crigler-Najjar like phenotype and Hemophilia B phenotype were observed.

Pending Application 08/927,165

Filed on 09/11/97

Inventors: Kmiec *et al.*

Our file: 7991-010-888

ATTORNEY DOCKET NUMBER: 7991-086-99

SERIAL NUMBER: 09/685,403

REFERENCE: **BF**

MAMMALIAN AND HUMAN REC2

This application claims benefit of the priority of U.S. provisional application Serial No. 60/025,929, filed September 11, 1996.

1. FIELD OF THE INVENTION

The present invention concerns the field of molecular genetics and medicine. Particularly, it concerns genes encoding a protein that is involved in homologous recombination and the repair of damaged genomic DNA in mammalian cells. Specifically the invention concerns the gene encoding a mammalian ATP-dependent homologous pairing protein; methods of using the gene to effect gene therapy; methods of using the gene and fragments of the gene to classify a mammalian tissue for medical purposes; and transgenic mice having had one or both alleles of the gene made inoperative. More specifically, the gene of the present invention is the human and murine homologs of the gene termed *REC2* previously isolated from the fungus *Ustilago maydis*.

2. BACKGROUND OF THE INVENTION

During the life of every organism the DNA of its cells is constantly subjected to chemical and physical events that cause alterations in its structure, i.e., potential mutations. These potential mutations are recognized by DNA repair enzymes found in the cell because of the mismatch between the strands of the DNA. To prevent the deleterious effects that would occur if these potential mutations became fixed, all organisms have a variety of mechanisms to repair DNA mismatches. In addition, higher animals have evolved mechanisms whereby cells having highly damaged DNA, undergo a process of programmed death ("apoptosis").

The association between defects in the DNA mismatch repair and apoptosis inducing pathways and the development, progression and response to treatment of oncologic disease is widely recognized, if incompletely understood, by medical scientists. Chung, D.C. & Rustgi, A.K., 1995, *Gastroenterology* 109:1685-99; Lowe, S.W., et al., 1994, *Science* 266:807-10. Therefore, there is a continuing need to identify and clone the genes that encode proteins involved in DNA repair and DNA mismatch monitoring.

Studies with bacteria, fungi and yeast have identified three genetically defined groups of genes involved in mismatch repair processes. The groups are termed, respectively, excision repair group, the error prone repair group and recombination repair group. Mutants in a gene of each group results in a characteristic phenotype. Mutants in the recombination repair group in yeast result in a phenotype having extreme sensitivity to ionizing radiation, a sporulation deficiency, and decreased or absent mitotic recombination. Petes, T.D., et al., 1991, in Broach, J.R., et al., eds., *THE MOLECULAR BIOLOGY OF THE YEAST SACCHAROMYCES*, pp. 407-522 (Cold Spring Harbor Press, 1991).

Several phylogenetically related genes have been identified in the recombination repair group: *recA*, in *E. Coli*, Radding, C.M., 1989, *Biochim. Biophys. Acta* 1008:131-145; *RAD51* in *S. cerevisiae*, Shinohara, A., 1992, *Cell* 69:457-470, Aboussekhra, A.R., et al., 1992, *Mol. Cell. Biol.* 12:3224-3234, Basile, G., et al., 1992, *Mol. Cell. Biol.* 12:3235-3246; *RAD57* in *S. cerevisiae*, *Gene* 105:139-140; *REC2* in *U. maydis*, Bauchwitz, R., & Holloman, W.K., 1990, *Gene* 96:285-288, Rubin, B.P., et al., 1994, *Mol. Cell. Biol.* 14:6287-6296. A third *S. cerevisiae* gene *DMC1*, is related to *recA*, although mutants of *DMC1* show defects in cell-cycle progression, recombination and meiosis, but not in recombination repair.

The phenotype of *REC2* defective *U. maydis* mutants is characterized by extreme sensitivity to ionizing radiation, defective mitotic recombination and interplasmid recombination, and an inability to complete meiosis. Holliday, R., 1967, *Mutational Research* 4:275-288. UmREC2, the *REC2* gene product of *U. maydis*, has been extensively studied. It is a 781 amino acid ATPase that, in the presence of ATP, catalyzes the pairing of homologous DNA strands in a wide variety of circumstances, e.g., UmREC2 catalyzes the formation of duplex DNA from denatured strands, strand exchange between duplex and single stranded homologous DNA and the formation of a nuclease resistant complex between identical strands. Kmiec, E.B., et al., 1994, *Mol. Cell. Biol.* 14:7163-7172. UmREC2 is unique in that it is the only eukaryotic ATPase that forms homolog pairs, an activity it shares with the *E. coli* enzyme *recA*.

U.S. patent application, Serial No. 08/373,134, filed January 17, 1995, by W.K. Holloman and E.B. Kmiec discloses *REC2* from *U. maydis*, methods of producing recombinant UmREC2 and methods of its use. Prior to the date of the present invention a

fragment of human REC2 cDNA was available from the IMAGE consortium, Lawrence Livermore National Laboratories, as plasmid p153195. Approximately 400 bp of the sequence of p153195 had been made publicly available on dbEST database.

The scientific publication entitled: ISOLATION OF HUMAN AND MOUSE GENES BASED ON HOMOLOGY TO REC2, July 1997, Proc. Natl. Acad. Sci. **94**, 7417-7422 by Michael C. Rice et al., discloses the sequences of murine and human Rec2, of the human REC2 cDNA. and discloses that irradiation increases the level of hsREC2 transcripts in primary human foreskin fibroblasts.

3. SUMMARY OF THE INVENTION

The invention provides nucleic acids encoding mammalian ATP-dependent homologous pairing proteins (a "mammalian recombinase") particularly, the human and murine ATP-dependent homologous pairing protein (hsREC2 and muREC2, respectively). The invention additionally provides DNA clones containing a copy of the mRNA encoding a mammalian recombinase (an "mREC cDNA") and DNA clones containing a copy of the genomic DNA containing an mREC gene or fragments thereof. In further embodiments, the invention concerns a nucleic acid comprising an mREC cDNA linked to a heterologous promoter, i.e., a promoter other than a mammalian recombinase promoter, so that a mammalian recombinase can be expressed or over-expressed in insect and mammalian cells and in bacteria. In one embodiment, the heterologous promoter is the polyhedrin promoter from the baculovirus *Autographica californica* and the invention provides for an isolated and purified mammalian recombinase, particularly isolated and purified hsREC2.

The invention provides several utilities of said nucleic acids and isolated and purified proteins. In the area of gene therapy and the construction of transgenic animals, the invention provides a method of enhancing homologous recombination between an exogenous nucleic acid and the genome of a cell by introducing a plasmid that expresses an mammalian recombinase into the cell, which method can be used to repair a genetic defect and thereby cure a genetic disease. Alternatively, for the construction of transgenic animals the invention provides a method of enhancing homologous recombination between an exogenous nucleic acid and the genome of a cell by introducing purified

mammalian recombinase into the cell. Alternatively, the invention provides for the construction of a transgenic animal, such as a mouse, having a transgenic mammalian recombinase gene operably linked to a strong promoter so that the recombinase is over expressed in some or all tissues. Such transgenic animals are highly adapted to undergo homologous recombination.

The invention additionally provides two methods of classifying a sample of human tissue for diagnosis and prognosis: by determining whether the cells of the sample contains two, one or no copies of *hsREC2*; and by determining whether the copies of *hsREC2* that said cells contain are mutated. For the purpose of diagnosis and classification of tissue samples the invention, firstly, provides paired oligonucleotides that are suitable for the PCR amplification of fragments of *hsREC2* and, secondly, identifies a microsatellite DNA sequence, D14S258, that is closely linked to *hsREC2*.

The invention further provides a transgenic mouse having one or both alleles of *muREC2* interrupted and thereby inactivated. The resultant transgenic animals, termed heterozygous and homozygous *REC2*-knock out mice, respectively, are susceptible to tumorigenesis by chemical carcinogens. *REC2*-knock-out mice can be used to determine whether there is a significant risk of carcinogenesis associated with a chemical or a process of interest. The reduced level or absence of *muREC2* makes *REC2*-knock-out mice a more sensitive test animal than wild-type.

The invention further provides a method of sensitizing target cells to DNA damage, such as from γ - or UV irradiation or from cytotoxic agents commonly used in oncologic therapy, which comprises causing the expression of high levels of recombinase in the target cell. The expression of such levels causes the cells to more readily undergo apoptosis in response to DNA damage. The invention yet further provides the *REC2* promoter a mammalian promoter that is inducible by irradiation or other DNA damaging agents.

4. BRIEF DESCRIPTION OF THE FIGURES

Figures 1A-1D.

Figures 1A and 1B show the derived amino acid sequence of *hsREC2* (SEQ ID NO:1) and the nucleic acid sequences of the *hsREC2* cDNA coding strand (SEQ ID NO:2), respectively. Figures 1C and 1D show the the derived amino acid

sequence of muREC2 (SEQ ID NO:3) and the nucleic acid sequences of the *hsREC2* cDNA coding strand (SEQ ID NO:4), respectively.

Figure 2A-2C. Figure 2A is an annotated amino acid sequence of *hsREC2*. Specifically noted are the nuclear localization sequence ("NLS"), A Box and B Box motif sequences, DNA binding sequence and a src-type phosphorylation site ("P"). Figure 2B is a cartoon of the annotated sequence, showing in particular the region 80-200 is most closely related to *recA*. Figure 2C shows the sequence homology between *hsREC2* and *Ustilago maydis* REC2. The region of greatest similarity, 43% homology, is in bold.

Figure 3. Figure 3 shows the DNA reannealing as a function of added baculovirus-produced hexahistidylHsREC2.

Figure 4. Figure 4 is a gel shift assay showing that binding of a *hsREC2*-thioredoxin fusion protein to ssDNA is ATP or γ -SATP dependent.

Figure 5A-5B. Figure 5A shows the frequency of repair of the Sickle Cell Disease mutation, as a function of added β^S - β^A chimeric repair vector (SC1), in the β -globin genes in a population of EBV-transformed human lymphoblasts in the presence or absence of the *hsREC2* expression vector pcHsREC2, pcDNA3 or pcCAT control plasmids or SC1 alone. Figure 5B shows the sequences of SC1, β^S and β^A in the region of the Sickle Cell mutation. Lower case a, c, g, and u indicates 2'-OMe nucleotides.

Figure 6. Figure 6 shows the reannealing of a 123 nt DNA fragment is catalyzed by GST/REC2 fusion protein.

Figures 7A1-2 and 7B1-2. Figures 7A1-2 and 7B1-2 show the sequence of the *hsREC2* and *muREC2* promoters respectively. The locations of sequences homologous to the sequences of known cis-acting radiation responsive elements in yeast are underlined and the corresponding yeast gene is indicated.

Figures 8A-8H. Figures 8A-8H show FACS histograms of Rnase treated, propidium iodide stained, CHO cells that have been transfected with either an *hsREC2* expressing plasmid (15C8) or an irrelevant control plasmid (Neo). The DNA content of the cells is displayed in the horizontal axis. The histograms are of unirradiated cells (8A, 8E) or of cells that are 24, 48 or 72 hours status post exposure to 15 J/m^2 UV irradiation (8B-8D, 8F-8H). The comparison shows that the expression of *hsREC2* increases the fraction of irradiated cells having less than the diploid DNA content, which is indicative

of apoptosis.

5. DETAILED DESCRIPTION OF THE INVENTION

As used herein, genes are *italicized*, e.g., *hsREC2*, while the corresponding protein is in normal typeface.

5.1 *hsREC2* AND THE STRUCTURE OF ITS PRODUCT *hsREC2*

The results of efforts to obtain *hsREC2* cDNA by hybridization under non-stringent conditions with *UmREC2* probes were unsatisfactory. Efforts were made to isolate a fragment of *hsREC2* by PCR amplification using primers that encode pentapeptides based on the *UmREC2* sequence. A mixture of four forward primers encoding residues 256-260 of *UmREC2*, GKTQM (SEQ ID NO:7), was constructed using inosine as the third base for the gly and thr codons and having a 5' noncoding GC dinucleotide, i.e., 5'-GC GGI AA(G/A) ACI CA(G/A) ATG-3'. A mixture of eight reverse primers complementary to the sequences that can encode residues 330-334 of *UmREC2*, YITSG, was synthesized, using inosine in the same way as the forward primers, i.e., 5'-CC ICC G(C/G)(T/A)¹ IGT IAT (A/G)TA-3'. The primers were used to amplify fragments of human genomic and cDNA libraries using the Expand™ system (Boehringer) coupled with two rounds of reamplification. After reamplification the fragments were cloned in pCRII (Invitrogen). Ten different mixtures of primers encoding a total of nine different pentapeptides were used and a total of about 60 fragments were sequenced. One 110 bp fragment from a human kidney cDNA library, *hsr110*, had statistically significant homology with *UmREC2*.

A computer search of the database dbEST was performed to find clones of cDNAs encoding proteins that have significant homologies with *UmREC2* and *hsr110*. The plasmid p153195 was identified as having significant homology with *UmREC2* and which contained *hsr110*. In one segment of 44 residues of *UmREC2* and *hsREC2*, there was 43% homology between *UmREC2* and *hsREC2*, i.e., 19 of the 44 residues of each sequence were identical. Additionally, there were 8 conservative substitutions. This

¹ Only two of the four combinations are complementary to ser codons, however, they are complementary to the ser codons most often used in humans.

region of high homology corresponds to residues 84-127 of *hsREC2* and residues 226-270 of *UmREC2*. See Figure 2C. Residues 226-270 of *UmREC2* is the portion of *UmREC2* that is most highly conserved when compared to *recA* and related members of the recombination repair group; cf. residues 40-95 of *recA*, 95-13 of *DMC1*, residues 100-144 of *RAD51*, and residues 160-204 of *RAD51*. See, e.g., Figure 2, Rubin et al., 1994 *supra*.

That clone p153195 lacked the 5' end of *hsREC2* was determined by the absence of an inframe start codon. The 5' end of *hsREC2* cDNA was obtained by PCR amplification of a cDNA library using a forward primer from the cloning vector and nested reverse primers based on p153195. An overlap of about 100 bp was identified which contained a unique restriction site that was used to reconstruct the full length *hsREC2* cDNA. The sequences of the reconstructed *hsREC2* cDNA and the derived sequence of *hsREC2* are given in Figure 1A-1B. The *hsREC2* cDNA encodes a protein of 350 amino acids, SEQ ID NO: 1 of Figure 1. The sequence of the *hsREC2* cDNA and its complement are SEQ ID NO: 2 and No: 3, respectively. The 5' boundary of p153195 was about nt 280 of SEQ ID NO: 2.

Comparisons of the *hsREC2* sequence with the *UmREC2* sequence reveals statistically significant, but distant homologies ($p = 2.8 \times 10^{-5}$). A similar level of homology is found between *hsREC2* and the yeast protein *DMC1*.

An expression vector containing the complete *hsREC2* cDNA under control of a strong promoter, for example, the cytomegalovirus promoter (pcHsREC2), can be constructed for over-expression of *hsREC2* in transfected eukaryotic cells. For the production of purified *hsREC2* a vector suitable for the expression of the *hsREC2* under control of the baculovirus polyhedrin promoter can be constructed. It is preferred to construct a vector that synthesizes a REC2 fusion protein consisting of a protein or peptide that aids in the purification of the product, such as a hexahistidyl peptide or glutathione S-transferase. Figures 1C and 1D show the derived amino acid and nucleic acid sequences of the murine REC2 (*muREC2*) cDNA

5.2 HOMOLOGS OF *hsREC2*

The present invention encompasses mammalian homologs of *hsREC2*. Nucleic

acids encoding the REC2 from any mammalian species can be identified and isolated by techniques, routine to those skilled in the art, using the sequence information of Figure 1A-1B and/or the *hsREC2* cDNA clone. Such routine techniques include use of the *hsREC2* cDNA or fragments thereof to probe cDNA and genomic libraries from other mammalian species and use of the sequence data to construct primers for PCR amplification of fragments of mammalian REC2 cDNA. The cloning of *hsREC2* and *muREC2* genomic DNA (gDNA) is described below.

High levels of transcripts of *hsREC2* can be found in heart and skeletal muscle, lung, pancreas, spleen and thymus, and placenta. Moderate or low levels of *hsREC2* transcripts are found in liver, kidney, brain and testes. Thus, the source of mRNA to construct cDNA libraries for obtaining mammalian REC2 clones is not critical. The sequence of residues 83-127, which corresponds to amino acids 226-270 of UmREC2, is particularly highly conserved and is, therefore, useful in identifying mammalian REC2 homologs.

Mammalian homologs of *hsREC2* can be identified by the presence of an amino acid sequence identity of greater than 80% and preferably greater than 90% compared to *hsREC2* in the highly conserved portions of the gene, i.e., the portion homologous to residues 83-127 of *hsREC2*. In a preferred embodiment the mammalian recombinase gene shares greater than 80% sequence identity with *hsREC2* gene within the about 130 bp segment that encodes the residues homologous with residues 83-127 of *hsREC2*. Such mammalian homologs of *hsREC2* will also have the above-noted activities of catalyzing DNA reannealing, ATPase activity and ATP-dependent ssDNA binding activity.

As used herein, a protein having each of these three activities is termed an ATP-dependent homologous pairing protein (a "mammalian recombinase"). A mammalian recombinase having greater than an 80% sequence identity with *hsREC2* is termed an "mREC2." Based on the extensive studies of bacterial and yeast homologous recombination proteins, those skilled in the art anticipate that all mammalian recombinases will have greater than 80% amino acid sequence identity with *hsREC2*, i.e., be an mREC2.

The invention further encompasses fusion proteins comprising a mammalian REC2 protein or fragment thereof, wherein the REC2 fragment displays at least one and

preferably each of the three above-noted activities to substantially the same extent as the native REC2. Those skilled in the art appreciate that the recombinant production and purification of mammalian proteins in bacterial and insect cell based expression systems is facilitated by the construction of fusion proteins that contain the protein of interest and a second protein that stabilizes the resultant fusion protein and facilitates its purification. Non-limiting examples of fusion proteins include hexahistidyl, Glutathione-S-transferase and thioredoxin fused to the amino terminus of REC2.

In one embodiment, the invention is a composition containing an isolated and purified protein, which is an ATP-dependent homologous pairing protein, i.e., is an ATP-dependent catalyst of DNA reannealing, is an ATPase, and binds ssDNA in the presence of ATP or γ -S-ATP, and which protein comprises a polypeptide of at least 115 amino acids which is substantially identical to a polypeptide found in a mammalian ATP-dependent homologous pairing enzyme. More preferably the isolated and purified protein comprises a polypeptide that is substantially identical to residues 80-200 of hsREC2. In a further embodiment, the isolated and purified protein of the invention comprises the polypeptide which is residues 2-350 of SEQ ID NO:1. As used herein, substantially identical means identical or having at most one conservative substitution per 20 amino acids. As used herein a human protein is an isolated and purified human protein if the composition containing the protein is substantially free of all other normally intracellular human proteins but a defined set of individually identified human proteins; similarly an isolated and purified mammalian protein is free of all other normally intracellular mammalian proteins except for a defined set of individually identified mammalian proteins. As used herein, "a composition which comprises a defined protein substantially free of a named material" means that the weight of the named material in the composition is less than 5% of the weight of the protein in the composition.

The invention further provides an isolated and purified nucleic acid derived from a mammalian species, i.e., derived from a cDNA or gDNA clone, that encodes a protein or fusion protein, having a sequence, which comprises the sequence of a mammalian ATP-dependent homologous pairing protein or a substantially identical sequence. As used herein, an isolated and purified nucleic acid is a nucleic acid isolated and purified free of nucleic acids encoding other mammalian proteins or fragments thereof. As used herein,

the sequence of a mammalian ATP-dependent homologous pairing protein means the sequence of a naturally occurring, i.e., wild-type ATP-dependent homologous pairing protein found in a mammal, or of any mutants of wild-type mammalian ATP-dependent homologous pairing protein. In preferred embodiments the nucleic acid of the invention encodes a protein that is greater than 80% sequence identical, or alternatively, more than 90% sequence identical to hsREC2. Those skilled in the art appreciate that the N-terminal and C-terminal one, two or three amino acids can be substituted or deleted without effect and, as used herein, are not considered a part of the sequence unless so specified. Those skilled in the art further appreciate that the insertion or deletion of one to four consecutive amino acids during the evolution of homologous proteins is common. Therefore, in the definition of sequence identity between proteins encompasses the introduction of as many as four, one to four residue gaps in one or both sequences to maximize identity.

The isolated and purified nucleic acids of the invention encompass not only cDNA and gDNA clones of mammalian genes encoding a mammalian ATP-dependent homologous pairing protein, but also nucleic acids derived from said cDNA and gDNA clones by site directed mutagenesis. By use of routine PCR techniques, those skilled in the art can make specific, predetermined changes in the sequence of a DNA. Site directed mutagenesis may be conducted by any method. The method of Ho, S.N., et al., GENE 77:51-59 (herewith incorporated by reference in its entirety), is suitable. According to the method of Ho, overlapping, mutated genome fragments are synthesized in two separate PCR reactions. Of the four primers are used in the two reactions, two are complementary to each other and introduce the desired mutation. The PCR reactions are performed so that the 3' end of the sense strand of one product is complementary to the 3' end of antisense strand of the other. The two PCR products are denatured, mixed and reannealed. The overlapping partial duplex molecules are then extended form a full length dsDNA, amplified in a third PCR reaction, the product isolated and inserted by conventional recombinant techniques into the parent gene. See, also, Liang, Q., et al., 1994, PCR Methods & Applic. 4:269-74; Weiner, M.P. & Costa, G.L., 1994, PCR Methods & Applic. 4:S131-136; Barrettino, D., et al., 1994, Nucleic Acids Research 22:541; Stemmer, W.P., et al., 1992, Biotechniques 13:214-220. By multiple

applications of such techniques any desired modifications in the sequence of a cloned DNA can be introduced. Thus, the nucleic acids of the invention are not limited to isolated and purified nucleic acids having naturally occurring sequences, but also include nucleic acids encoding a ATP-dependent homologous pairing protein having substantially the same sequence as a naturally occurring mammalian recombinase.

The compositions of the invention further include compositions comprising not only mammalian recombinases isolated and purified free of mammalian proteins, but also compositions comprising any isolated and purified ATP-dependent homologous pairing protein having substantially the same sequence as a naturally occurring mammalian recombinase.

The hsREC2 sequence contains several sequences that have been identified with specific functions in other proteins. Figure 2A shows the sequence of hsREC2 and indicates the locations of nuclear localization sequence, four sequences associated with recA, namely A box, B box, a src-like phosphorylation site and a DNA binding site. Those skilled in the art will appreciate that, as was found for UmREC2, not all portions of a mREC2 protein are essential for the *in vitro* activities that characterize ATP-dependent homologous binding proteins. However, the region of about 120 amino acids from about residue 80 to residue 200, which is recA-like, is essential for these activities.

5.3 THE USE OF mREC2 AND mREC2 ENCODING GENES TO EFFECT HOMOLOGOUS RECOMBINATION BETWEEN THE GENOME OF A CELL AND AN EXOGENOUS NUCLEIC ACID

In one embodiment of the invention, a plasmid that expresses an mREC2 is used to increase the rate of homologous recombination between an exogenous nucleic acid and the genome of a cell. In one embodiment, the exogenous nucleic acid is a chimeric repair vector (CRV), which is an oligonucleotide having mixed ribo- and deoxyribonucleotides. The structure of CRV are disclosed in U.S. patent applications Serial No. 08/353,657, filed December 4, 1994, and Serial No. 08/664,487, filed June 17, 1996, which are hereby incorporated by reference in its entirety. U.S. application Serial No. 08/640,517, entitled "Methods and Compounds for Curing Diseases Caused by Mutations," filed May 1, 1996, by E.B. Kmiec, A. Cole-Strauss and K. Yoon, (the '517 Application), which is hereby incorporated by reference in its entirety, describes the use

of CRV to repair mutations that cause diseases. Particularly, the '517 Application concerns the repair of mutations that affect hematopoietic cells such as the mutation in β -globin that causes Sickle Cell Disease.

According to the present invention, the cell having a disease-causing mutation to be repaired (the target cell) is removed from the subject. The target cells are then transfected with a nucleic acid having a promoter operably linked to a nucleic acid encoding a mREC2 (an mREC2 expression vector) such that a mammalian ATP-dependent homologous pairing protein is over-expressed in the target cell. For most types of human cells, the immediate early promoter from cytomegalovirus is suitable. Because the persistent over-expression of a mammalian ATP-dependent homologous pairing protein can effect the growth and differentiation of the target cell, the mREC2 expression vector should be incapable of replication in the target cell. The mREC2 expression vector can be introduced into the target cell by any technique known to those in the field or to be developed. Liposomal compositions such as LIPOFECTIN^(TM) and DOTAP^(TM) are suitable.

After transfection with the mREC2 expression vector, the target cells are cultured for twenty four hours and then a CRV designed to repair the disease causing mutation is introduced into the target cells, according to the methods of the '517 Application, and repaired target cells are then reimplanted into the subject. Alternatively, the repaired target cells can be frozen and reimplanted at a clinically opportune time.

Figure 5 shows the results of the use of an mREC2 expression vector to enhance the effectiveness of a CRV that repairs the mutation that causes Sickle Cell Disease in a human EB-transformed lymphoblastoid cell line. These data show that at a concentration of CRV of about 100 ng/ml, the pretreatment of the target cells with the mREC2 expression vector pcHsREC2, labelled "pchREC2" in Figure 5, caused an about 5 fold increase, from 12% to 65%, in the percent of repaired copies of β -globin. At 250 ng/ml, over 80% of the copies of β -globin were repaired. At higher concentrations of CRV, the differences between pcHsREC2 treated target cells and control target cells become less marked.

The present invention is exemplified by the use of a non-replicating episome to introduce an mREC2 cDNA gene (*hsREC2*), operably linked to a cytomegalovirus (CMV) promoter, into the target cell and to transiently express mREC2. Alternative embodiments

of the invention can be produced by introducing the copy of a genomic gene, which can be linked to the homologous mREC2 promoter or, alternatively, modified so that the homologous promoter is replaced by a CMV or other heterologous promoter. Further variants that can be used to increase homologous recombination in different situations include linkage of either mREC2 cDNA or gDNA to tissue specific promoters such as a CD4, immunoglobulin, insulin or globin promoter. By use of tissue specific promoters, transgenic animals, particularly mice, rats and swine can be constructed that overexpress mREC2 in only one particular tissue. In yet a further alternative embodiment the promoter can be an inducible promoter. An inducible promoter particularly suitable for the present invention is a tetracycline inducible promoter, which is described in U.S. Patent No. 5,464,758, which is incorporated by reference in its entirety.

Those skilled in the art will further appreciate that an mREC2 encoding gene can be constructed that contains some but not all introns of the complete mREC2 gDNA. Such a gene is a mixture of mREC2 gDNA and mREC2 cDNA fragments. As used herein the term "an mREC2 gene" is to be understood to denote, generically, mREC2 cDNA, mREC2 gDNA or a nucleic acid encoding a full length REC2 protein comprising mREC2 gDNA and mREC2 cDNA fragments.

The present invention further encompasses the use of mREC2 expression vectors to facilitate the construction of transgenic animals using cultured embryonic stem cells ("ES cells") according to the method of Capecchi, M.R., 1989, Science 244: 1288 and U.S. Patent 5,487,992, Col. 23-24, which are incorporated by reference in their entirety. A transgenic mouse having a inducible mREC2 gene introduced can be constructed. ES cells from such a transgenic mouse can be obtained and induced to have elevated levels of mREC2. Such cells will more readily undergo homologous recombination with a chimeric mutational vector ("CMutV"), an oligonucleotide having a similar structure and function to those of CRV, that can be used to introduce specific mutations into targeted wild-type genes. By use of CMutV, second and higher generation transgenic animals having further targeted genetic alterations can be constructed.

A further embodiment of the invention concerns the use of isolated and purified mREC2 protein in the construction of transgenic animals. Those skilled in the art of constructing transgenic animals understand that transgenic animals are constructed by

direct injection of a nucleic acid into the pronucleus of an ovum according to the method described Brinster, R.L. et al., 1989, PROC. NATL. ACAD. SCI 86:7087; see also U.S. Patent No. 4,873,191 to T.E. Wagner and P.C. Hoppe, which are hereby incorporated by reference in their entirety. Such direct injection results in the random integration of the injected nucleic acid. As noted above techniques for the introduction of transgenes by homologous recombination have been developed, however, such techniques require a specialized embryonic stem cell line, which is available only for mice, and, in addition require that the genetic alteration be designed so that homologous recombinants can be selected in culture, since the rate of homologous recombination is very low.

Because the use of the present invention in conjunction with CMutV permits a specific alteration to be introduced into a large fraction, e.g., 80%, of the copies of a target gene, those skilled in the art will appreciate that the invention provides a practical technique for the construction of transgenic animals wherein the function of both alleles of a specifically targeted gene has been deleted ("knocked-out") by homologous recombination using ova directly injected with a REC2 CMutV mixture.

Transgenic animals are constructed according to the invention by injecting a ova pronucleii with mREC2 protein and the CMutV. In a preferred embodiment a mixture of the CMutV and a mREC2 protein is injected into the ova pronucleus. In a preferred embodiment the nucleic acid to be injected is a CMutV that introduces a stop codon or a frameshift mutation into the gene to be knocked out. The concentration of protein to be used is about one molecule of mREC2 protein per between 5,000 base pairs and 50 base pairs of the CMutV, preferably one molecule of mREC2 protein per about 100-500 base pairs of the CMutV. Alternatively, the CMutV can be replaced by a conventional linearized DNA fragment containing homologous regions flanking a mutator region.

5.4 THE CONSTRUCTION OF *mu*REC2-KNOCK-OUT MICE

The invention additionally provides transgenic mice that contain inactivated *mu*REC2. Such heterozygous *mu*REC2-knock-out transgenic mice can be constructed by injection of a murine embryonic blastocyst with an embryonic stem cell line (ES cells) that has the appropriate mutation in *mu*REC2 (*mu*REC2^{ko}). The technique of Nichols, J., et al., 1990, DEVELOPMENT 110:1341-48 can be used. Further teaching regarding the

construction of transgenic mice using embryonic stem cell-injected blastocysts can be found in U.S. Patent No. 5,487,992 to Capecchi and Thomas, which is hereby incorporated by reference in its entirety. Homozygous *muREC2*-knock-out mice can be obtained by the intercross of heterozygous *muREC2*-knock-out mice and selection of offspring that are homozygous for the *muREC2*^{ko} allele.

Without limitation, a *muREC2*^{ko} gene can be made in two ways. A CMutV can be constructed according to U.S. patent No. 5,565,350, which is designed to introduce one or more stop codons at different positions within *muREC2* (an "*muREC2*^{ko} chimeric vector"). ES cells line can be treated with the *muREC2*^{ko} chimeric vector. Preferably several *muREC2*^{ko} chimeric vectors, designed to introduce redundant stop codons are used to reduce the reversion rate. After treatment, the ES cells can be cloned and the loss of a functional *muREC2* gene confirmed by sequence analysis or by PCR amplification using primers specific for the mutated codons.

Alternatively, a dicistronic targeting construct can be used to introduce a *muREC2*^{ko} mutation. Mountford, P.; et al., 1994, Proc. Natl. Acad. Sci. 91:4303-07. More specifically, targeting vector is constructed having a cassette consisting of, in 5' to 3' order, a splice acceptor site, the 500 bp internal ribosome entry site (IRES) from encephalomyocarditis virus (EMCV), a fusion gene β geo, that has both β -galactosidase and G418 resistance activity, and an polyadenylation signal from SV40. In the targeting construct, the cassette is inserted, as an example without limitation, between two fragments from the introns 3' and 5' of the second exon of the *muREC2* gene, wherein the 5' most exon is the first exon, the exon immediately 3' to the 5'-most exon is the second exon etc. The length of the fragments can be preferably between about 500 bp and 5,000 bp.

The linearized targeting construct can be introduced into an ES cells by any technique suitable for the transfection of DNA into ES cells. The *muREC2* gene of the transfected ES cells undergoes homologous recombination whereby the cassette replaces the second exon such that the cassette is transcribed from the *muREC2* promoter and the β geo protein is translated by ribosomes bound to the IRES. ES cells having the cassette integrated into transcriptionally active genes can be selected by exposing the transfected cells to G418 and by histochemical staining to detect galactosidase positive cells.

Typically as many as 70% to 90% of $\beta\text{gal}^+/\text{neo}^r$ double transformants have undergone homologous recombination of the targeted gene.

Homozygous *muREC2*^{ko} mice have an increased susceptibility to mutation caused by chemical and physical agents. Such animals can be used to determine if products are mutagenic and more specifically if such products are carcinogens. Both homozygous and heterozygous *muREC2*^{ko} mice will also be more susceptible to the development of benign and malignant tumors. These animals can be used to originate tumors of different tissue types for use in biomedical studies.

5.5 THE CLASSIFICATION OF SAMPLES OF HUMAN TISSUE BY EXAMINATION OF THE *hsREC2* GENES OF THE SAMPLE

Those skilled in the art appreciate that there is a close connection between the a cell's capacity to remove chemically induced mutations and replication errors from its DNA and the cell's potential to develop the genetic changes that result in the development and progression of malignancies. Altonen, L.A., 1993, Science 260:812-816; Chung, D.C., & Rustgi, A.K., 1995, Gastroenterology 109:1685-99. A cell's capacity to remove mutations and replication errors can be classified by determining, firstly, whether the cell contains the normal, i.e., diploid number of copies of a gene that is essential for DNA mismatch repair and, secondly, by determining whether the copies that are present have been altered, i.e., contain mutations. Cells having a diminished capacity to remove DNA mismatches because of defects in their REC2 are malignant or are more likely to become malignant due to the further accumulation of mutations.

In one embodiment, the invention consists of classifying a human tissue according to the number of copies of the *hsREC2* gene per diploid genome. The reduction of the number to less than two indicates that some cells of the tissue can have a reduced capacity to repair DNA mismatches, because a mutation in the remaining copy would cause the absence of ATP-dependent homologous pairing activity. The number of copies of a gene can be readily determined by quantitative genomic blotting using probes constructed from labelled nucleic acids containing sequences that are fragment of SEQ ID NO:2 or a complement thereof. An alternative method of determining the number of *hsREC2* genes per diploid genome in a sample of tissue relies on the fact that the *hsREC2* gene is located in bands 14q23-24 and, particularly, that it is tightly linked to the

proximal side of the marker D14S258 and also tightly linked to the marker D14S251. The loss of a copy of a *hsREC2* gene in an individual who is heterozygous at a locus linked to the D14S258 marker can be inferred from the loss of the heterozygosity.

An alternative embodiment of the invention consists of classifying a sample of human tissue according to whether or not it contains an unmutated copy of a *hsREC2* gene. The *hsREC2* gene of the sample and the *hsREC2* of a standard tissue can be compared by any technique known to those skilled in the art or to be developed. A sensitive technique suitable for the practice of this embodiment of the invention is single strand conformational polymorphism (SSCP). Orita, M., et al., 1989, *Genomics* 5:874-879; Hayashi, K., 1991, *PCR Methods and Applic.* 1:34-38. The technique consists of amplifying a fragment of the gene of interest by PCR; denaturing the fragment and electrophoresing the two denatured single strands under non-denaturing conditions. The single strands assume a complex sequence-dependent intrastrand secondary structure that affects the strands electrophoretic mobility. Therefore comparison of an amplified fragment of a *hsREC2* gene from a sample of tissue with the amplified fragment from a *hsREC2* gene of a standard tissue is a sensitive technique for detecting mutations in the *hsREC2* of the sample.

The absence of a copy of an unmutated *hsREC2* gene in a sample of tissue indicates that the cells of the tissue have undergone or likely will undergo transformation into a malignant phenotype.

In a further alternative embodiment of the tissue sample can be classified by Southern blotting of the DNA of the sample. The presence of tissue specific bands in the blot is evidence that at least one copy of the *REC2* gene of the sample has undergone a mutational event. In yet a further embodiment of the invention, the tissue sample can be classified by amplifying a fragment of the *REC2* gene, by PCR, and analyzing the fragment by sequencing or by electrophoresis to determine if the sequence and length of the amplified fragment is that which can be expected from a normal *REC2* gene.

Without limitation, particular types of tissue samples that can be classified according to the invention include tumors which are associated with cytogenetic abnormalities at bands 14q23-24. Such tumor types include renal cell carcinomas and ovarian cancers Mittelman, F., 1994, *Catalog of Chromosome Aberrations in Cancer*,

(Johansson, B. and Mertens, F. eds.) Wiley-Liss, New York, pp 2303-2484. Also suitable for classification according to the method of the invention are tumor types that show a loss of heterozygosity of markers linked to the region 14q23-24. Such tumor types include meningiomas, neuroblastomas, astrocytomas and colon adenomas. Cox, D.W., 1994, Cytogenetic Cell Genet. 66:2-9. Of particular interest is the high rate of breast adenocarcinomas that have been found to have either mutated *hsREC2* genes or to have lost heterozygosity of the microsatellite DNA at the closely linked locus D14S258.

In addition to the above described methods the embodiments of the invention include a kit comprising a pair of oligonucleotides suitable for use as primers to amplify a fragment of a *hsREC2* gene, which pair consists of a 5'-primer having a sequence of a fragment of SEQ ID NO:2 and a 3'-primer having a sequence of a fragment of its complement wherein the 3'-primer is complementary to a portion of the sequence of SEQ ID NO:2 that lies 3' of the location of the 5'-primer sequence. The length of the 3' and 5'-primers is at least 12 nucleotides and preferably between about 16- and 25-nucleotides and more preferably between 18 and 24 nucleotides. The invention further consists of oligonucleotides having a sequence of a fragment of SEQ ID NO:2 or its complement and a label, which are suitable for hybridization with genomic blots of the *hsREC2* gene. Labels include radiolabels such as ³²P, fluorescent labels or any label known or to be developed that allows for the specific detection of a nucleic acid sequence.

The plasmid pcHsREC2, in which the *hsREC2* cDNA is operably linked to a CMV immediate early promoter has been deposited on August 20, 1996, in the ATCC, Rockville, MD, and accorded accession No. 97685. The plasmid was deposited under the name "pcHuREC2," but is referred to herein as pcHsREC2 for consistency. The plasmid pcHsREC2 is derived from commercially available plasmid pcDNA3 (Invitrogen, Inc.) and contains a 1.2 Kb insert that encodes *hsREC2*, which can be removed from pcHsREC2 by cutting with the restriction enzymes XbaI and KpnI.

EMBL-3-type λ phage clones, designated λ 5A and λ 1C, which contain a 12 Kb and 16Kb fragment of the 5' and 3' region of the *hsREC2* gene, respectively, were deposited on August 20, 1996, as accession No. 97683 and No. 97682, respectively.

AFIXII type λ phage clones, designated λ 5D2a and λ 7B1a, which contain a 14 Kb

and 14.9Kb fragment of the 5' and 3' region of the *muREC2* gene of strain 129SVJ, respectively, were deposited on August 22, 1996 and August 20, 1996, as accession No. 97686 and No. 97684, respectively. The inserts of λ 5D2a and λ 7B1a are released by cutting with a NotI restriction enzyme.

5.6 THE REC2 PROMOTER

The promoters of *hsREC2* and *muREC2* were cloned. The *hs REC2* promoter was cloned by a two step PCR-based promoter walking technique. Briefly, blunt ended genomic fragments are made by digestion with *DraI* and *SspI*, in the first and second step respectively. The restriction fragments are ligated to adapters. A primary PCR amplification is performed using a gene specific primer from the 5' extreme of the gene and an adapter specific primer. A secondary PCR is performed using nested, gene and adapter specific primers. The first step, primary and secondary gene specific primers were 5-CAG ACG GTC ACA CAG CTC TTG TGA TAA-3' (SEQ ID NO:8) and 5'-ACC CAC TCG TTT TAG TTT CTT GCT AC-3 (SEQ ID NO:9), respectively. The second step promoter walking primary and secondary primers were 5'-TAG AGA GAG AGA GAG AGC GAG ACA G-3' (SEQ ID NO:10) and 5'-GTC GAC CAC GCG TGC CCT ATA G-3' (SEQ ID NO:11), respectively. The first step and second step fragments were 0.8 and 0.9 Kb in length respectively.

The *mu REC2* promoter was sequenced by digestion of the clone λ 5D2a with *XbaI*. The promoter was found on the largest fragment, of about 7 Kb. The sequences of the *hsREC2* and *muREC2* are given in Figures 7A and 7B respectively.

The level of *REC2* transcripts in cultured human foreskin fibroblasts had been shown to be increased when the cells were exposed to ^{137}Cs irradiation. Several yeast genes have been identified that are radiation inducible and the radiation sensitive cis-acting control sequences from the promoters of such genes have been identified. See references cited in footnotes to Tables I-III. The sequence of the *hsREC2* and *muREC2* promoters were therefore inspected for the presence of such sequences. Figure 7A and 7B demonstrates that numerous such sequences were present. Tables I-III show the sequence of the yeast UV responsive elements, their positions in the yeast gene in which they are found and the reference to the scientific publication where they are

described.

The radiation induceability of the hsREC2 gene was directly assayed using UV radiation and the luciferase reporter gene in transiently transfected HeLa cells. The hsREC2 promoter was operably linked to a luciferase reporter gene and to the SV40 enhancer, placed downstream of the poly A addition signal. Any strong enhancer can be used, e.g., the enhancer from Cytomegalovirus, Hepatitis B Virus, α -fetoprotein, Rous Sarcoma Virus or Simian Virus 40. In this construct hsREC2 promoter was, in the absence of radiation approximately as strong a promoter as the SV40 immediate early promoter. When the cells were UV irradiated (35 J/m^2 UV) the hsREC2 promoter showed an approximate two to three fold increase in activity. See Section 6.8, below.

A radiation induceable promoter can be used to increase the susceptibility of cells to radiation as, for example, in conjunction with radiation therapy of a cancer. A construct containing a hsREC2 promoter operably linked to a "suicide gene", e.g., herpes thymidine kinase, can be introduced into mitotically active cells using a retrovirus based vector. A tumor can be irradiated and, simultaneously, gancyclovir, a DNA antimetabolite prodrug that is converted by herpes thymidine kinase, can be administered.

Those skilled in the art appreciate that the activity of the REC2 promoter can be further localized by testing the activity of the fragment after deletions having been made. A functional, radiation induceable promoter that is smaller than the fragment of Figure 7A or 7B can be found. Accordingly as used herein a human REC2 promoter and a murine REC2 promoter is defined as a DNA having the sequence found in Figure 7A or 7B, respectively, or a fragment thereof, wherein said fragment is a promoter in HeLa cells. The terms hsREC2 promoter and muREC2 promoter refer to DNA molecules having the sequences found in Figure 7A and 7B respectively. A REC2 promoter from any species can be defined analogously. Accordingly, in one embodiment, the invention concerns a composition containing a only a defined number of types of DNA molecules, one of which molecules comprises a REC2 promoter. As used herein such composition is said to comprise an isolated and purified REC2 promoter. In an alternative embodiment, the invention concerns a plasmid having a bacterial origin of replication (henceforth a "cloning plasmid"), which plasmid comprises a mammalian REC2 promoter and

specifically a human or a murine REC2 promoter. Those skilled in the art will further appreciate that the cis-acting radiation sensitive control elements present in the sequences of Figure 7A and 7B can be identified by systematic testing of fragments having the appropriate deletions. Accordingly, there can be REC2 promoters, as defined above, that are less radiation inducible than the hsREC2 promoter. As used herein a mammalian REC2 promoter is said to be radiation inducible if the promoter shows at least a two fold increase in activity and a REC2 promoter is termed "three fold inducible" if it shows a three fold increase when tested under the conditions wherein hsREC2 gives at least a four fold increase.

In further embodiments the REC2 promoter is operably linked to an enhancer. The present invention is illustrated by use of the SV40 enhancer. Those skilled in the art appreciate that any enhancer that is as strong as the SV40 enhancer can be used. Alternative enhancers include Cytomegalovirus, Hepatitis B Virus, α -fetoprotein, Rous Sarcoma Virus or Simian Virus 40 enhancers.

Table I UASs of <i>Saccharomyces cerevisiae</i> DNA repair genes					
Gene	Location	Sequences	SEQ ID NO	References	
PHR1	-103	CGAGGAAGCAGT	15	13, 14	
	-110	CGAGGAAGAAAA	16		
RAD2	-166	GGAGGCATTAAA	17	5 15, 16	
RAD23	-295	GGTGGCGAAATT	18		
RAD51	-215	CGTTACCCTAT	19		
RAD54	-256	CGTTACCCAAT			
Consensus		GGAGGARRNANA	20		
		C T C			

Table II. UASs of *Saccharomyces cerevisiae* DNA repair genes

Gene	Location	Sequences	SEQ ID NO.	References
Rhp51+	-290	CGTT_CCCTAT	21	11
	-260	CCTA_CCCTAA	22	
RAD51	-215	CGTTACCCTAT	23	12
RAD54	-256	CGTTACCCAAT	24	17
RNR3	-429	CGGTTGCCATG	25	18
Consensus		CGTTACCCTAT	26	

Table III URSSs of *Saccharomyces cerevisiae* DNA repair genes

Gene	Position	Sequences	SEQ ID NO	References
MAG	-215	GTAGGTCGAA	27	1
PHR1	-103	CGAGGAAGCA	28	2
	-109	CGAGGAAGAA	29	2
RAD2	-169	CGTGGAGGCA	30	1,2,3,4,5
RAD51	-157	CGTGGTGGGA	31	6,12
DDR48	-271	CGAGGATGAC	32	1,7
	-322	CGTGGTTGAT	33	1,7
RNR2	-374	CGAGGTCGCA	34	8,9
RNR3	-467	CTAGGTAGCA	35	1,10
rhp51+	-233	GTAGGTGTTA	36	11
	-213	CTAGGTAACA	37	11
RAD16	-309	CATGGTTGCC	38	1
Consensus		CGTGGTNGAA	39	1
		A A CC		

References to Tables I-III:

1. Xiao W. et. al., 1993, Mol. Cell. Biol., **13**, 7213-7221
2. Sebastian. J. et. al., 1990, Mol. Cell. Biol., **10**, 4630-4637
3. Madura, K. et. al., 1986, J. Bacteriol., **166**, 914-923
4. Reynolds, P. et. al., 1985, EMBO J., **4**, 3549-3552
5. Siede, W. et. al., 1989, Mol. Microbiol., **3**, 1697-1707
6. Basile, G. et. al., 1992, Mol. Cell. Biol., **12**, 3235-3246
7. Treger, J. M. et. al., 1990, Mol. Cell. Biol., **10**, 3174-3184
8. Elledge, S. J. et. al., 1989, Mol. Cell. Biol., **9**, 5373-5386
9. Hurd, H. K. et. al., 1989, Mol. Cell. Biol., **9**, 5359-5372
10. Yagle, K. et. al., 1990, Mol. Cell. Biol., **10**, 5553-5557
11. Jang, Y. K. et. al., May 23, 1996, Molecular & General Genetics, **251** (2), 167-175,
12. Aboussekhra A. et. al, 1992, Mol. Cell. Biol., **12**, 3224-3234
13. Sancar, G. B. ,1985, Nucleic Acids Research, **13**, 8231-8246
14. Sancar, G. B. et. al., 1995, Nucleic Acids Research, **23**, 4320-4328
15. Jones, J. S. et. al., 1991, Nucleic Acids Research, **19**, 893-898
16. Watkins, J. F. et. al., 1993, Mol. Cell. Biol., **13**, 7757-7765
17. Cole, G. M. et. al., 1989, Mol. Cell. Biol., **9**, 3314-3322
18. Elledge S. J. et. al. ,1989a, Mol. Cell. Biol., **7**, 4932-4940

REC2-TRANSFECTANTS ARE SENSITIZED TO IRRADIATION

One embodiment of the present invention is a plasmid or other isolated purified DNA molecule in which a mREC2 cDNA is operably linked to a strong promoter, which is preferably a constitutive promoter, e.g., a CMV immediate early promoter. In a further embodiment the invention consists of a mammalian cell that is transfected with such plasmid or isolated purified DNA and which over expresses Rec2. The overexpression of Rec2 causes a mammalian cells to be hypersensitive to DNA damaging agents such as alkylating agents, e.g., cyclophosphamide, γ-ray or UV-irradiation.

Accordingly, the present invention can be used to sensitize a set of cells that can be selectively transfected with a Rec2 expressing plasmid. Such sensitization can be used in conjunction with conventional oncologic chemotherapy or irradiation therapy to treat malignant disease.

6. EXAMPLES**6.1**

**The production of recombinant hsREC2 protein
by baculovirus infection of *Autographica
californica***

To facilitate the construction of an *hsREC2* expression vector, restriction sites for XhoI and KpnI were appended by PCR amplification to a the *hsREC2* cDNA. The *hsREC2* cDNA starting at nt 71 was amplified using the forward primer 5'-GAG CTCGAG GGTACC C ATG GGT AGC AAG AAA C-3' (SEQ ID NO:14), which placed the XhoI and KpnI sites (underlined) 5' of the start codon. The recombinant molecule containing the entire coding sequence of *hsREC2* cDNA, can be removed using either XhoI or KpnI and the unique XbaI site located between nt 1270 and 1280 of SEQ ID NO:2.

A vector, pBacGSTSV, for the expression of HsREC2 in baculovirus infected *Spodoptera frugiperda* (Sf-9) insect cells (ATCC cell line No. CRL1711, Rockville MD), was obtained from Dr. Zailin Yu (Baculovirus Expression Laboratory, Thomas Jefferson University). The vector pVLGS was constructed by the insertion of a fragment encoding a *Schistosoma japonicum* glutathione S-transferase polypeptide and a thrombin cleavage site from pGEX-2T (described in Smith & Johnson, GENE 67:31 (1988)), which is hereby incorporated by reference, into the vector into the vector pVL1393. A polyA termination

signal sequence was inserted into pVLGS to yield pBacGSTSV. A plasmid containing the 1.2 Kb *hsREC2* fragment was cut with KpnI, the 3' unpaired ends removed with T4 polymerase and the product cut with XbaI. The resultant fragment was inserted into a SmaI, XbaI cut pBacGSTSV vector to yield pGST/*hsREC2*.

Recombinant virus containing the insert from pGST/*hsREC2* were isolated in the usual way and Sf-9 cells were infected. Sf-9 cells are grown in SF900II SFM (Gibco/BRL Cat # 10902) or TNM-FH (Gibco/BRL Cat # 11605-011) plus 10% FBS. After between 3-5 days of culture the infected cells are collected, washed in Ca^{++} and Mg^{++} free PBS and sonicated in 5ml of PBS plus proteinase inhibitors (ICN Cat # 158837), 1% NP-40, 250 mM NaCl per 5×10^7 cells. The lysate is cleared by centrifugation at 30,000 xg for 20 minutes. The supernatant is then applied to 0.5 ml of glutathione-agarose resin (Sigma Chem. Co. Cat # G4510) per 5×10^7 cells. The resin is washed in a buffer of 50 mM Tris-HCl, pH 8.0, 150 mM NaCl and 2.5 mM CaCl_2 , and the *hsREC2* released by treatment with thrombin (Sigma Chem. Co. Cat # T7513) for 2 hours at 23°C in the same buffer. For certain experiments the thrombin is removed by the technique of Thompson and Davie, 1971, *Biochim Biophys Acta* 250:210, using an aminocaproyl-p-chlorobenzylamide affinity column (Sigma Chem. Co. Cat # A9527).

6.2 Detection of the Enzymatic Properties of *hsREC2* protein

Baculovirus produced hexahistidyl*hsREC2* was tested in a DNA reannealing assay as described in Kmiec, E.B., & Holloman, W.K., 1982, *Cell* 29:367-74. The results, Figure 3, showed that *hsREC2* catalyzes the reannealing of denatured DNA. An optimal reaction occurred at about 1 *hsREC2* per 50-100 nucleotides.

Further studies to characterize *hsREC2* showed that it catalyzes the reaction $\text{ATP} \rightarrow \text{ADP} + \text{PO}_4$. Similar to *recA*, at ATP concentrations of $< 100 \mu\text{M}$, there is cooperativity between *hsREC2* molecules; the Hill coefficient (1.8) suggests that the functional unit for ATP hydrolysis is at least a dimer. Gel retardation experiments were performed to determine the ATP dependence of *hsREC2* binding to ssDNA. The results of these experiments showed that *hsREC2* binds ssDNA only in the presence of ATP or its non-hydrolyzable thio analog γ -SATP. Figure 4. Again the *hsREC2* results parallel those of *recA*.

Further examples of specific assays using isolated and purified hsRec2 are as follows:

6.2.1 Binding to Single Stranded DNA

A 73 nucleotide single stranded DNA (SS) was ^{32}P end labelled using polynucleotide kinase. DNA binding was carried out using 0.25 ng of labeled SS in 25 mM Tris-HCl, pH 7.4, 10 mM MgCl_2 , 4 mM ATP, and 1 mM DTT and protein. hsRec2-thioredoxin was partially purified on a ThiobondTM column (Invitrogen) and desalted/concentrated using a Microcon 30 spin column (Amicon). Approximately 0.3 μg protein was added. The reaction mixture was incubated 30 min. at 37°C, following which sucrose was added to facilitate loading onto a polyacrylamide gel. The mixture was loaded onto a 12% nondenaturing gel in 90 mM Tris, 90 mM borate, pH 8.3, 2 mM EDTA for 3 hours at 150 V. The gel was then dried and exposed overnight. Approximately 3% of the label was retarded in the presence of ATP or $\gamma\text{S-ADP}$, while reduced amounts of label were bound in the absence of either of ATP or $\gamma\text{S-ADP}$.

6.2.2 Catalysis of Reannealing of DNA

Reannealing of a 123 nucleotide fragment was determined as follows. The single stranded 123 nucleotide (SS) was ^{32}P end labelled using polynucleotide kinase. Varying amounts of affinity purified GST-hsRec2 fusion protein was added to 0.5 ng of SS in 25 μl of 20mM TrisHCl pH 7.5, 10 mM MgCl_2 , 0.5 mM DTT with 5 mM ATP optionally present. Samples were incubated 30 min. at 37°C, followed by phenol/chloroform extraction to stop the reaction, followed by a second 30 min. incubation at 37°C. The reaction mixture was then electrophoresed as in section 6.2.1, above, and autoradiographed. The results, shown in Figure 6, demonstrate that GST-hsREC2 catalyzes the reannealing of the SS in both the presence and absence of ATP.

6.3 Overexpression of *hsREC2* Suppresses UVC-Induced Mutation

To determine whether the presence of hsRec2 protects cultured cells from UVC induced mutation a CHO cell line was transfected with a mixture of linearized pcHsREC2 and pCMVneo and a clone resistant to G418 was selected ("15C8 hsREC2"). Elevated levels of hsREC2 expression were confirmed by immunoblotting using rabbit antisera raised to baculovirus produced hsRec2 fusion proteins.

Mutability was determined as follows. 1.6×10^6 15C8 hsREC2 cells were plated in a 100 mm petri dish and exposed to 0 or between 2.0 and 5.0 J/m² UV radiation. After 7 days of culture, the remaining cells were exposed to 40 μ M 6-TG. Surviving cells had undergone an inactivation of the HPRT gene. After a further 7-10 days of culture the number of colonies was counted. The mutation frequency was adjusted for the cloning efficiency of the population which was determined by plating a limiting number of cells without 6-TG.

The results showed that the untransfected, pCMVneo and 15C8 hsREC2 cells had mutation rates of 1.7, 6.2 and 0.4 per million, respectively, without UVC irradiation. After UVC radiation the mutation rates observed were, in three experiments, between 94 and 16, 61 and 74, and 3 and 37, per million, for untransfected, pCMV transfected and 15C8 hsREC2 cells, respectively. Thus, the expression of *hsREC2* caused a marked decrease in the susceptibility of CHO cells to UVC induced mutation as well as a drop in the spontaneous mutation frequency.

6.4 **Enhanced Repair of β -globin in Cultured, EB-transformed Human Lymphoblasts**

SC1, a chimeric vector designed to repair the mutation found in Sickle Cell Disease β -globin, contained two blocks of ten 2'-O-methyl RNA residues each, flanking an intervening block of five DNA residues, see Figure 5B. When the molecule was folded into the duplex conformation, one strand contained only DNA residues while the other strand contained the RNA/DNA blocks. In this case, the internal sequence is complementary to the β^S globin sequence over a stretch of 25 residues that span the site of the β^S mutation, with the exception of a single base (T) which is in bold and designated with an asterisk. The five DNA residues flanked by RNA residues were centered about the mutant T residue in the β^S coding sequence. Genomic sequences of the β^A , β^S , and closely-related δ -globin genes are also displayed in Figure 3 with the specific site of β^S mutation printed in bold.

Lymphoblastoid cells were prepared as follows. Heparin-treated blood was obtained from discarded clinical material of a patient with sickle cell disease. Mononuclear cells were prepared from blood (≈ 8 ml) by density gradient centrifugation in Ficoll and

infected with Epstein-Barr virus which had been propagated in the marmoset cell line B95-8 (Coriell Institute for Medical Research #GM07404D). Infections were performed with addition of 0.1 mg leucoagglutinin PHA-L in 10 ml RPMI medium supplemented with 20% fetal bovine serum in a T25 flask. Cultures were fed twice a week starting on day 5 and were considered established once 60-70% of the cells remained viable at day 21. The β^A and β^S lymphoblastoid cells were maintained in RPMI medium containing 10% fetal bovine serum.

The EBV-transformed lymphoblastoid cells were transiently transfected with either the vector pcDNA3 or the vector having inserted *hsREC2* cDNA (pCHsREC2). Transfection was done using mixtures of 15 μ l DOTAP and 2.5 μ g DNA, as detailed below. After transfection the cells were incubated for 24 hours and then treated with varying amounts of SC1.

SC1 was introduced into the above-described lymphoblastoid cells homozygous for the β^S allele as follows. Cells (1×10^5 per ml) were seeded in 1 ml of medium in each well of a 24-well tissue culture plate the day prior to the experiment. Transfections were performed by mixing chimeric oligonucleotides in amounts ranging from 0 to 250 ng, with 3 μ l of DOTAP (N-[1-(2,3-Dioleoyloxy)propyl]-N,N,N-trimethylammonium methylsulfate, Boehringer-Mannheim) in 20 ml of 20 mM HEPES, pH 7.3, incubated at room temperature for 15 min, and added to the cultured cells. After 6 h the cells were harvested by centrifugation, washed and prepared for PCR amplification following the procedure of E.S. Kawasaki, PCR Protocols, Eds. M.A. Innis, D.H. Gelfand, J.J. Sninsky and T.J. White, pp146-152, Academic Press, (1990).

Correction of the single base mutation was assessed by taking advantage of well known restriction fragment length polymorphisms resulting from the β^S mutation, R.F. Greeves et al., 1981, Proc. Natl. Acad. Sci. 78:5081; J.C. Chang and Y.W. Kan, 1982, N. Eng. J. Med. 307:30; S.H. Orkin et al., *ibid.*, p. 32; J.T. Wilson et al., 1982, Proc. Natl. Acad. Sci. 79:3628. The A to T transversion in the β^S allele results in the loss of a Bsu36I restriction site (CCTGAGG). Thus, the β^S allele can be detected by Southern hybridization analysis of genomic DNA cut with Bsu36I. A 1.2 Kb Bsu36I DNA fragment of the β -globin gene present normally is absent in the β^S allele and is replaced by a diagnostic 1.4 Kb fragment. When genomic DNA recovered from homozygous β^S

lymphoblastoid cells was analyzed by this procedure, the expected 1.4 Kb fragment was observed. However, two fragments were observed in DNA from cells transfected with the SC1 CRV. The presence of the 1.2 Kb fragment in addition to the 1.4 Kb fragment indicates partial correction of the β^s allele had taken place in a dose-dependent fashion.

The results of the experiment are shown in Figure 5A. At 100 ng and 250 ng of SC1 between 65% and 85% of the β^s alleles were mutated to β^A alleles in the cells pre-transfected with pcHsREC2, compared to between 10% and 25% in the non pre-transfected cells and negligible levels in the control transfected cells. At levels of SC1 between 25 ng and 50 ng of SC1, no mutations were detected in any of the control cell populations while between 30% and 40% of the β^s alleles were mutated to β^A alleles in the cells pre-transfected with pcHsREC2.

These results show that the over expression of *hsREC2* causes marked increase in the susceptibility of a cell to mutation by a chimeric mutation vector such as SC1.

6.5 Identification and Isolation of mREC2 gDNA Clones

Genomic blots of human and murine, strain 129 SVJ, DNA were made using XbaI and BamHI digests. Following transfer to Zeta-Probe™ membranes (Bio-Rad) the membranes were prehybridized for 30' at 55°C in 0.25M NaHPO₄, pH7.2, 7% SDS, 1 mM EDTA and hybridized overnight with a random primed full length HsREC2 probe. Wash was 2X for 20' at 42°C in 0.04M NaHPO₄, pH7.2, 5% SDS, 1 mM EDTA and 1X each at 42°C and 50°C for 20' in 0.04M NaHPO₄, pH7.2, 1% SDS, 1 mM EDTA. The results were bands of the following sizes: Human-XbaI 6.0, 4.1, 2.6, 2.0 and 1.5 Kb; Human-BamHI 9.5, 8.5, 6.5, 4.6, 1.5 Kb; Murine-XbaI 9.0, 6.0, 4.1, 3.5, 1.9, 0.8 Kb; and Murine-BamHI 8.0, 2.7 and 1.8.

To identify and propagate clones containing mREC2 from cDNA or DNA libraries standard techniques for cloning were employed using λ -phage libraries. A human genome library in EMBL-3 and a murine genomic library in λ FIXII were screened. Phage plaques were transferred to hybridization filters by standard techniques and the filters were probed with radiolabelled *hsREC2* cDNA. After hybridization the filters were washed. A wash consisting of twice at 42°C for 20' in 2x SSC, 0.1% SDS followed by thrice at 50°C for 20' in the same solution was used to isolate murine gDNA clones. To

isolate human gDNA clones a the wash procedure was: twice 20 min. at 42°C in 40 mM NaHPO₄, pH 7.2, 1 mM EDTA, and 5% SDS; followed by once for 20 min. at 50°C in the same solution except for 1% SDS.

The 5' and 3' fragments of *muREC2* and *hsREC2* gDNA were recovered in the following λ phage clones: λ 5D2a (14 Kb insert, 5' *muREC2*); λ 7B1a (14.9 Kb insert, 3' *muREC2*); λ 5A (12 Kb insert, 5' *hsREC2*); λ 1C (16 Kb insert, 3' *hsREC2*), each of which has been deposited in the ATCC, Bethesda, MD.

Fragments of genomic clones can be used as probes of genomic blots to identify rearrangements deletions or other abnormalities of *hsREC2* in tumor cells. Those skilled in the art further appreciate that by routine sequence analysis and comparison with the sequence of SEQ ID NO: 2, the boundaries of the exons and introns of *hsREC2* can be identified. Knowing the sequence of at the intron/exon boundaries allows for the construction of PCR suitable for the amplification and analysis of each exon as alternatives to the methods of section 6.6.

6.6 Elevated Incidence of Abnormalities in *hsREC2* in Adenocarcinomas of the Breast

Samples of 30 primary ductal carcinoma of the breast were analyzed by Southern blot, probed with the *hsREC2* cDNA and by a high resolution gel of the PCR product of the microsatellite marker D14S258, which is closely linked to the *hsREC2* gene. Ten of the thirty samples gave abnormal results in one of the two assays and 3 showed abnormalities by both assays. In contrast none of 16 samples of primary renal cell carcinoma showed clear abnormalities in a Southern blot.

6.6.1 Loss of Heterozygosity of Microsatellite DNA Linked to *hsREC2*

The location of *hsREC2* was found to be tightly linked to the proximal side of the microsatellite marker D14S258. Because there is extensive polymorphism in the lengths of microsatellite sequences most individuals are heterozygous at the D14S258 locus. Primers specific for unique sequences flanking the polymorphic locus can be used to generate PCR fragments whose length is allele specific. Primers specific for D14S258 were obtained from the Dr. Lincoln Stein, Whitehead Institute, MIT, Cambridge MA. The "5'" primer is 5'-TCACTGCATCTGGAAGCAC-3' (SEQ ID NO:12) and the "3'" primer is

5'-CTAACTAAATGGCGAGCATTGAG-3' (SEQ ID NO:13). PCR was performed with a genomic DNA concentration of 2.0 ng/ μ l, a primer concentration of 10.0 μ M, 10.0 μ M dNTP, 500 μ M Tris HCl, pH 9.2, 17.5 μ M MgCl₂, 160 μ M (NH₄)₂SO₄, and a polymerase concentration of 0.03 U/ μ l. Amplification was performed for 35 cycles of 50 seconds each, alternating between 57°C and 94°C, followed by an extension of 7 minutes at 72°C and preceded by an initial heat soak of 5 minutes at 94°C. The expected product is about 160-170 nucleotides in length.

A comparison of the products of PCR amplification of tumor and normal tissue control DNA using the flanking primers can reveal the loss of either or both D14S258 loci, which suggests that the linked *hsREC2* has also been lost.

The results of analysis 7 of 30 samples breast tumors showed a complete or partial loss of one allele at locus D14S258.

These results show that instability and loss of a genetic locus tightly linked to the location of *hsREC2* is found in a large fraction of human ductal adenocarcinoma of the breast.

6.6.2 Frequent Rearrangements of *hsREC2*

Genomic DNA from samples of 16 primary renal and 30 primary breast tumors tumor tissue were digested with either XbaI or BamHI restriction enzymes, electrophoresed in a 0.8% agarose gels and processed for hybridization with labeled random primed copies made from the *hsREC2* cDNA. After transfer, Zetaprobe™ blotting membranes were UV crosslinked, prehybridized at 65°C for 20 min in 0.25M NaHPO₄, pH 7.4, 7% SDS, 1mM EDTA and then hybridized overnight under the same conditions. The membranes were pre-washed once with 40mM NaHPO₄, pH 7.2, 5% SDS, 1mM EDTA at 42°C for 20 min, then washed repeatedly at 60°C in the same solution, except for 1% SDS, until background levels were achieved in the periphery of the membrane. The filters were then exposed to film.

Six of the 30 examples of carcinoma of the breast showed rearrangements or abnormalities while none of the 16 samples of renal cell carcinoma showed clear rearrangements.

6.7 Construction of a MuREC2^{ko} containing ES cell line

The muREC2 gDNA clone λ 5D2a contains the first two exons. The second exon is located on 3.6 Kb Eco R1 fragment, approximately 1.2 Kb from the fragment's 5' border. The second exon contains a unique *Stu*I site into which was inserted the IRES- β geo poly A cassette, Mountford, P., et al., 1994, Proc. Natl. Acad. Sci. **91**, 4303-4307. ES cells were cultured on primary mouse embryo fibroblasts according to standard protocols, Hogan, B., et al., 1996, MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Press. Approximately 2×10^7 ES cells were transfected by electroporation with 25 μ g linearized DNA. Selection was begun at 36 hours and continued until day 8 with 250 μ g/ml G418. Thirty colonies were isolated and tested by *Xba*I digest and Southern blot; one colony was found to lack the wild type size *Xba*I fragment and to have a novel fragment of the predicted size. Transgenic mice are constructed from this ES cell line by conventional techniques. *Ibid*.

6.8 The HsREC2 Promoter Is Radiation Induceable

A 1.8 Kb fragment immediately 5' to the hsREC2 start codon was cloned. The fragment was tested as a promoter using the luciferase reported gene construct, pGL2, (Promega Cat. No. E1611), luciferase activity was measured using the luciferase reported test kit (Boehringer Mannheim Cat. No. 1669 893).

The activity of the promoter is assayed in HeLa cells as follows. The HeLa cells are trypsinized on day -1 and plated at 6.6×10^5 / 60 mm well in 3.0 ml of DMEM. On day two at -1h the medium is replaced with serum free medium and the cells are transfected with various quantities of the plasmid with DOSPER at a DNA:DOSPER ratio of 1:4. At 5 hour an additional 3.0 ml of medium supplemented with FBS is added; at 24 hours the cells are irradiated with UV light (Stratalinker). Cells are harvested at 48 hours and proteins extracted and assayed. Control experiments done with the same plasmid having the SV40 immediate early promoter in place of the hsREC2 promoter.

UV Irradiation (Joules/meter ²)	DNA Added (Micrograms)			
	3 μ g	2.4 μ g	1.2 μ g	0.6 μ g
0 J/m ²	655.6 ¹	494 ²	27.5	32.8
15 J/m ²	951.5	1287	28.9	28.7
25 J/m ²	1033.6	1398	35.8	44.2
35 J/m ²	1134.6	1786	84.89	68.4

1. The corresponding luciferase is 513.9 pSV40-luc-SV40 enhancer at 0 Joules/meter².
2. The corresponding luciferase is 384 pSV40-luc-SV40 enhancer at 0 Joules/meter².

When the 3' 0.8 Kb of the hsREC2 promoter was tested beginning with nt 869 of SEQ ID NO: 5, it was determined that this 0.8 Kb fragment contains a promoter having reduced activity but which is also shows an about 5 fold induceability with 35 J/m² UV radiation in HCT 116, which cell line contains a normal p53 gene. The preferred form of the REC2 induceable promoter in HCT 116 is the shortened form starting at nt 869.

6.9 The Expression of REC2 Causes Increased Radiation Sensitivity

UV irradiation induces apoptosis in stable transfectants expressing wild-type HsRec2 but not truncated or full length with an altered tyrosine 163 site. In order to measure the effects of REC2 expression on the rate of UV induced radiation CHO cells were irradiated. During the 24 hour long recovery period following irradiation, more CHO cells expressing wild-type HsRec2 were observed to die than the control cells that expressed an irrelevant or nonfunctional proteins. To determine whether cell death was a result of apoptosis, asynchronous cells were irradiated at a dose of 15 J/m², and fixed in ethanol at 24, 48 and 72 hours following irradiation. FACS analysis was conducted as follows: Cells were trypsinized, washed once with PBS and fixed in 70% ethanol at least 30 minutes at 4°C Cell pellets were treated with DNase-free Rnase for 30 minutes at 70°C at a final concentration of 0.16 mg/ml and stained in propidium iodide (0.05 mg/ml)

for 15 minutes, then stored overnight prior to analysis by FACS. The FACS analysis and determination of the percentage of cells in G1, S and G2 phases (Multicycle Flow program) was carried out in the Cell Cycle Center at the Kimmel Cancer Institute of Thomas Jefferson University. Cells from duplicate cultures were harvested at the same time points, and frozen at -80°C. for DNA isolation. DNA was isolated using a QIAGEN Blood Kit (QIAGEN Inc., Chatsworth, CA) and stored at 4°C. until run on gels. DNA was run on 1% agarose gels in TAE buffer and stained 30 minutes with a 1:10,000 dilution of SYBR Green I (FMC, Rockland, ME). Gels were then scanned using a FluorImager (Molecular Diagnostics, San Diego, CA).

Four cell types were used for analysis; CHO cells containing the empty vector (Neo^r), CHO cells expressing HsRec2Δ103-350 (3D2), HsRec2^{ala63} (PH4), and the wild-type HsRec2 (15C8). A sub-G1 population was detected at 24, 48, and 72 hours following irradiation for CHO cells expressing the wild-type HsRec2 only. To confirm that apoptosis was occurring, DNA was isolated from cells, and run on a 1% agarose gel, stained with SYBR Green I and scanned. For each time interval compared, 15C8 exhibited a more pronounced ladder than the other clones. Although there appears to be a small amount of apoptosis for the clone expressing HsRec2^{ala63} it is considerably lower than for the wild-type HsRec2 clone, and neither the Neo^r or the transfectants expressing the truncated protein are comparable. Therefore, the G1 delay and apoptosis require the wild-type HsRec2, and suggests that perhaps cooperation between a mutant p53 present in CHO cells and Rec2 may be responsible for genome surveillance in these cells.

The results of the FACS analysis of the HsRec2 expressing and the Neo^r expressing clones are given in Figures 8A-8H.

SEQUENCE LISTING

(1) GENERAL INFORMATION

- (i) APPLICANT: Kmiec, Eric B.
Holloman, William K.
Rice, Michael C.
Smith, Sheryl T.
Shu, Zhigang

- (ii) TITLE OF THE INVENTION: Mammalian and Human Rec2

- (iii) NUMBER OF SEQUENCES: 39

- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Kimeragen, Inc.
 - (B) STREET: 300 Pheasant Run
 - (C) CITY: Newtown
 - (D) STATE: PA
 - (E) COUNTRY: USA
 - (F) ZIP: 18940

- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Diskette
 - (B) COMPUTER: IBM Compatible
 - (C) OPERATING SYSTEM: DOS
 - (D) SOFTWARE: FastSEQ for Windows Version 2.0

- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:

- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Hansburg, Daniel
- (B) REGISTRATION NUMBER: 36156
- (C) REFERENCE/DOCKET NUMBER: 7991-010-999

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: 215-504-4444
- (B) TELEFAX: 215-504-4545
- (C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 350 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Met	Gly	Ser	Lys	Lys	Leu	Lys	Arg	Val	Gly	Leu	Ser	Gln	Glu	Leu	Cys
1				5					10					15	
Asp	Arg	Leu	Ser	Arg	His	Gln	Ile	Leu	Thr	Cys	Gln	Asp	Phe	Leu	Cys
			20					25						30	
Leu	Ser	Pro	Leu	Glu	Leu	Met	Lys	Val	Thr	Gly	Leu	Ser	Tyr	Arg	Gly
			35				40					45			
Val	His	Glu	Leu	Leu	Cys	Met	Val	Ser	Arg	Ala	Cys	Ala	Pro	Lys	Met
			50				55					60			
Gln	Thr	Ala	Tyr	Gly	Ile	Lys	Ala	Gln	Arg	Ser	Ala	Asp	Phe	Ser	Pro
65					70				75					80	
Ala	Phe	Leu	Ser	Thr	Thr	Leu	Ser	Ala	Leu	Asp	Glu	Ala	Leu	His	Gly
				85					90					95	
Gly	Val	Ala	Cys	Gly	Ser	Leu	Thr	Glu	Ile	Thr	Gly	Pro	Pro	Gly	Cys
			100						105					110	
Gly	Lys	Thr	Gln	Phe	Cys	Ile	Met	Met	Ser	Ile	Leu	Ala	Thr	Leu	Pro
			115						120					125	
Thr	Asn	Met	Gly	Gly	Leu	Glu	Gly	Ala	Val	Val	Tyr	Ile	Asp	Thr	Glu
			130					135						140	

CCGTCTGAGT	AGACATCAGA	TCCTTACCTG	TCAGGACTTT	TTATGTCTTT	CCCCACTGGA	180
GCTTATGAAG	GTGACTGGTC	TGAGTTATCG	AGGTGTCCAT	GAACTTCTAT	GTATGGTCAG	240
CAGGGCCTGT	CCCCCAAAGA	TGCAAACGGC	TTATGGGATA	AAAGCACAAA	GGTCTGCTGA	300
TTTCTCACCA	GCATTCTTAT	CTACTACCCT	TTCTGCTTTG	GACGAAGCCC	TGCATGGTGG	360
TGTGGCTTGT	GGATCCCTCA	CAGAGATTAC	AGGTCCACCA	GGTTGTGGAA	AAACTCAGTT	420
TTGTATAATG	ATGAGCATTT	TGGCTACATT	ACCCACCAAC	ATGGGAGGAT	TAGAAGGAGC	480
TGTGGTGTAC	ATTGACACAG	AGTCTGCATT	TAGTGCTGAA	AGACTGGTTG	AAATAGCAGA	540
ATCCCGTTTT	CCCAGATATT	TTAACTACTGA	AGAAAAGTTA	CTTTTGACAA	GTAGTAAAGT	600
TCATCTTTAT	CGGGAACTCA	CCTGTGATGA	AGTTCACAA	AGGATTGAAT	CTTTGGAAGA	660
AGAAATTATC	TCAAAAGGAA	TTAACTTGT	GATTCTTGAC	TCTGTTGCTT	CTGTGGTCAG	720
AAAGGAGTTT	GATGCACAAC	TTCAAGGCAA	TCTCAAAGAA	AGAAACAAGT	TCTTGGCAAG	780
AGAGGCATCC	TCCTTGAAGT	ATTTGGCTGA	GGAGTTTTCA	ATCCCAGTTA	TCTTGACGAA	840
TCAGATTACA	ACCCATCTGA	GTGGAGCCCT	GGCTTCTCAG	GCAGACCTGG	TGTCTCCAGC	900
TGATGATTTG	TCCCTGTCTG	AAGGCACTTC	TGGATCCAGC	TGTGTGATAG	CCGCACTAGG	960
AAATACCTGG	AGTCACAGTG	TGAATACCCG	GCTGATCCTC	CAGTACCTTG	ATTGAGAGAG	1020
AAGACAGATT	CTTATTGCCA	AGTCCCCTCT	GGCTCCCTTC	ACCTCATTTG	TCTACACCAT	1080
CAAGGAGGAA	GGCCTGGTTC	TTCAAGCCTA	TGGAAATTCC	TAGAGACAGA	TAAATGTGCA	1140
AACCTGTTCA	TCTTGCCAAG	AAAAATCCGC	TTTTCTGCCA	CAGAAACAAA	ATATTGGGAA	1200
AGAGTCTTGT	GGTGAAACAC	CCATCGTTCT	CTGCTAAAAC	ATTTGGTTGC	TACTGTGTAG	1260
ACTCAGCTTA	AGTCATGGAA	TTCTAGAGGA	TGTATCTCAC	AAGTAGGATC	AAGAACAAGC	1320
CCAACAGTAA	TCTGCATCAT	AAGCTGATTT	GATACCATGG	CACTGACAAT	GGGCACTGAT	1380
TTGATAACAT	GGCACTGACA	ATGGGCACAC	AGGGAACAGG	AAATGGGAAT	GAGAGCAAGG	1440
GTTGGGTTGT	GTTTCGTGGAA	CACATAGGTT	TTTTTTTTTA	ACTTTCTCTT	TCTAAAATAT	1500
TTCATTTTGA	TGGAGGTGAA	ATTTATATAA	GATGAAATTA	ACCATTTTAA	AGTAAACAAT	1560
TCCGTGGCAA	CTAGATATCA	TGATGTGCAA	CCAGCATCTC	TGTCTAGTTC	CCAAATATTT	1620
CATCACCCCC	AAAAGCAAGA	CCCATAACCA	TTATGCAAGT	GTTTCCTATTT	CCCCCTCCTC	1680
CCAGCTCCTG	GGAAACCACC	AATCTACTTT	TTTTCTATGG	CTTTACCTAA	TCTGGAAATT	1740
TCAAATAAAT	GGGATCAAAT	AGTTTCCCAA	AAAAAAAAAA	AAAAAAAAAA	AAAAAAA	1797

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 350 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Ser Ser Lys Lys Leu Arg Arg Val Gly Leu Ser Pro Glu Leu Cys

1	5	10	15
Asp Arg Leu Ser Arg Tyr Leu Ile Val Asn Cys Gln His Phe Leu Ser			
20	25	30	
Leu Ser Pro Leu Glu Leu Met Lys Val Thr Gly Leu Ser Tyr Arg Gly			
35	40	45	
Val His Glu Leu Leu His Thr Val Ser Lys Ala Cys Ala Pro Gln Met			
50	55	60	
Gln Thr Ala Tyr Glu Leu Lys Thr Arg Arg Ser Ala His Leu Ser Pro			
65	70	75	80
Ala Phe Leu Ser Thr Thr Leu Cys Ala Leu Asp Glu Ala Leu His Gly			
85	90	95	
Gly Val Pro Cys Gly Ser Leu Thr Glu Ile Thr Gly Pro Pro Gly Cys			
100	105	110	
Gly Lys Thr Gln Phe Cys Ile Met Met Ser Val Leu Ala Thr Leu Pro			
115	120	125	
Thr Ser Leu Gly Gly Leu Glu Gly Ala Val Val Tyr Ile Asp Thr Glu			
130	135	140	
Ser Ala Phe Thr Ala Glu Arg Leu Val Glu Ile Ala Glu Ser Arg Phe			
145	150	155	160
Pro Gln Tyr Phe Asn Thr Glu Glu Lys Leu Leu Leu Thr Ser Ser Arg			
165	170	175	
Val His Leu Cys Arg Glu Leu Thr Cys Glu Gly Leu Leu Gln Arg Leu			
180	185	190	
Glu Ser Leu Glu Glu Glu Ile Ile Ser Lys Gly Val Lys Leu Val Ile			
195	200	205	
Val Asp Ser Ile Ala Ser Val Val Arg Lys Glu Phe Asp Pro Lys Leu			
210	215	220	
Gln Gly Asn Ile Lys Glu Arg Asn Lys Phe Leu Gly Lys Gly Ala Ser			
225	230	235	240
Leu Leu Lys Tyr Leu Ala Gly Glu Phe Ser Ile Pro Val Ile Leu Thr			
245	250	255	
Asn Gln Ile Thr Thr His Leu Ser Gly Ala Leu Pro Ser Gln Ala Asp			
260	265	270	
Leu Val Ser Pro Ala Asp Asp Leu Ser Leu Ser Glu Gly Thr Ser Gly			
275	280	285	
Ser Ser Cys Leu Val Ala Ala Leu Gly Asn Thr Trp Gly His Cys Val			
290	295	300	
Asn Thr Arg Leu Ile Leu Gln Tyr Leu Asp Ser Glu Arg Arg Gln Ile			
305	310	315	320
Leu Ile Ala Lys Ser Pro Leu Ala Ala Phe Thr Ser Phe Val Tyr Thr			
325	330	335	

Ile Lys Gly Glu Gly Leu Val Leu Gln Gly His Glu Arg Pro
340 345 350

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1525 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GGGAGCCCTG GAAACATGAG CAGCAAGAAA CTAAGACGAG TGGGTTTATC TCCAGAGCTG	60
TGTGACCGTT TAAGCAGATA CCTGATTGTT AACTGTCAGC ACTTTTAAAG TCTCTCCCCA	120
CTAGAACTTA TGAAAGTGAC TGGCCTGAGT TACAGAGGTG TCCACGAGCT TCTTCATACA	180
GTAAGCAAGG CCTGTGCCCC GCAGATGCAA ACGGCTTATG AGTTAAAGAC ACGAAGGTCT	240
GCACATCTCT CACCGGCATT CCTGTCTACT ACCCTGTGCG CCTTGGATGA AGCATTGCAC	300
GGTGGTGTGC CTTGTGGATC TCTCACAGAG ATTACAGGTC CACCAGGTTG CGGAAAAACT	360
CAGTTTTGCA TAATGATGAG TGTCTTAGCT ACATTACCTA CCAGCCTGGG AGGATTAGAA	420
GGGGCTGTGG TCTACATCGA CACAGAGTCT GCATTTACTG CTGAGAGACT GGTGAGATT	480
GCGGAATCTC GTTTTCCACA ATATTTTAAAC ACTGAGGAAA AATTGCTTCT GACCAGCAGT	540
AGAGTTCATC TTTGCCGAGA GCTCACCTGT GAGGGGCTTC TACAAAGGCT TGAGTCTTTG	600
GAGGAAGAGA TCATTTTCGAA AGGAGTTAAG CTTGTGATTG TTGACTCCAT TGCTTCTGTG	660
GTCAGAAAGG AGTTTGACCC GAAGCTTCAA GGCAACATCA AAGAAAGGAA CAAGTTCTTG	720
GGCAAAGGAG CGTCCTTACT GAAGTACCTG GCAGGGGAGT TTTCAATCCC AGTTATCTTG	780
ACGAATCAAA TTACGACCCA TCTGAGTGGA GCCCTCCCTT CTCAAGCAGA CCTGGTGTCT	840
CCAGCTGATG ATTTGTCCCT GTCTGAAGGC ACTTCTGGAT CCAGCTGTTT GGTAGCTGCA	900
CTAGGAAACA CATGGGGTCA CTGTGTGAAC ACCCGGCTGA TTCTCCAGTA CCTTGATTCA	960
GAGAGAAGGC AGATTCTCAT TGCCAAGTCT CCTCTGGCTG CCTTCACCTC CTTTGTCTAC	1020
ACCATCAAGG GGAAGGCCT GGTTCCTTCAA GGCCACGAAA GACCATAGGG ATACTGTGAC	1080
CTTTGTCTAG TGCTGATTGC ATGTGACTCA TGAAATGAAA CAGGACTGCG CTGCTTGGAA	1140
AAAGGAAACG GAAGCCAACA TAATGAGGAT TAATTGGTTG GTTGCTGTTG AGGTGGTAAC	1200
AGTGATTTC AACCCGGAAG GTGAAGATGA AGAAGCCTTT ATCCAGTCTC TGGATGCAGA	1260
GGCTAGGGGC TCCACCACCG TGGGATGTCA GCGGCCATCG TAATAATTTG CACTTACACA	1320
AGCACCTTTC AGCCATGCCC CTCAAAGTGG TTCAGCCACA TTAATTAATT AAAGCCCACA	1380
ATCCCCCTAG GGAGAGCAGG AGGGGGACTA ACAAGATTTG TAATTACAGA AGGGAAAAAT	1440
TCCGAATAAA GTATTGTTCC GCCAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	1500
AAAAAAAAAA AAAAAAAAAA AAAAA	1525

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1699 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CGACGGCCCG GGCTGGTATT ATAGCAGGTA TCACTTGGTT TTCTACTGGG GGAAACAAGT	60
CATTGCTAAC AAATTCCCAT GGGAGAGAAA TGAGGAGGAT GTATTTTTGT TTGTGAGAGG	120
TGTGTATGTA TGTATATTGT GTGTGCGTGT GTGTGTGTGT GAGAGAGAGA GATTGATTCA	180
GTCTGATTCA GAGAATTTAG GTGTTAAATA GAAATTTGGG CCATGGTATT GGAAATAAAC	240
AAATATATAC ATTCTCAGTA TACATATATT TTCATTCCAA AATGTTACTT CTTTTCTGAT	300
AACTATATTG CTTTATTCCC TTGGATCCAT GAAGAGTTCC TGTTTCAGTT CGTTCAGAG	360
GATACTTCTT TACCATCTCA ATGAGATATA CAGCTTCTCC TTTGTATGCA TTAAGAGACT	420
CACAGTAATT CTTTTTTAGC TCTGTGAAGA TAAATCTTTC ATGAGCCTCA TTTACCCCTA	480
GCAAAGTACA ATAGTGAAAT TTAAGTGCAT GTGAGAATAT AAGCAGCTAG TGTAATAAAG	540
AACATTTTGG GCCAGGTCTG ATCGCTCATG CCTGTAATCC CAGCACTTTA GGAGGTCAAG	600
GCGAGAGGAT CACTTGAGCC CAGGAGTTCG AGACCAGCTT GGGCAACATG GCAAAACCCT	660
GTCTCTACAA AAAATACAAA AATTGGGCAG GCATGGTGTC GACCCAGTCT CTACAAAAAA	720
TACAAAAATT AGCCAGACAT GGTGGTGCAC GCTTGTGGTC CCAGCTACTT GGGAGGCTGA	780
GGTAGGAGGA TTGCTTGAGC CCAGGAGGGG GAGGTTGCAG TGAGCTGAGA TCGAGCCACT	840
GCACTCCAGC TGGGGTGACA GAGCCAGACC TGTCTCGCTC TCTCTCTCTC TCTATATATA	900
TATATTTAAA AAGAACATTT TAATACTGCA GTGATAAAAT CTCATTTGAT TCAGAAGGTG	960
TGCTCTGACT CCTAGAAAAA GGAAGAGTCA AATATGATTA TGGACTTGCA GTAGAGTGTA	1020
ATGGTTAAGA GGATAGGTTT CAGAATTAGA CTGCCTGGAT TCAAATTCTG GATCAGTTAT	1080
TTATGGTTTC TGGTGACAAT GGAAGTACTA ACTTTTCCAG GCTTTAGTTT TCTCATATGT	1140
AAAAAAGGGG CCAATAATCT ACTTTCCTTC TAGGGCTATT GAGAAGATTA AATGTGATAA	1200
TTTAGATAAG TTTTGGAACA GTGCCTGGTA TGTGGTAGGT GCTCCATAAA TATACCTATT	1260
GCCGTTACAG TGCAATGTAA ATTGTTACAG TGCAATAGAC TTTCTAGTAG TTCTGTTTGG	1320
AAATATGCCT TGAAAGTTAA TTACATTTCC AAATAAAATT TATACATGCA TTGGAACATT	1380
TTAAGATGCT CTACAAATGT GAAGTGGTAC TATATTCATG TAGTAAATAT CAATTAATTG	1440
TGTGAAATTA TATTTGAGGT TGCCTTGTAG ATTTTCTATG TGCCTGTTTG ACGAACAATT	1500
GTCCCTCCTA TTTAAAACAT TAAAAAGGT TCTATAGCAT TCCTTTATCA GTAATATTTT	1560
TAACACAATA TGTTTCATTT TGCATATGGA GAACTTGAG GAATTTTTAA TTTTGTTTTG	1620
GATAGCCTAT TCACTATCAC TTATGTTATA TTCTGTTGTT TTTTTCATGG TTCTTCTTTT	1680
CTTTGCTGGA TCTGGAGGC	1699

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2147 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TATCTCAGTA	GCACGTGCAC	ATAGCAACTA	CAATACCTGT	CACATAAATG	TAGTTACTTG	60
AATATATGTC	TCTTCATTCT	TCAATTGTAA	GTATGCAAAA	GGGAGGACAT	AAGCTTAGCA	120
TAGCATGTGC	TTAATATTGG	TGAAAGAAAC	AAATGAATAG	AGAATGTTAT	ATTTGGAGAG	180
TTTATATTAT	ATTTGGGAGA	GTAGGGAAAA	AACTTGAAGC	CATAAGCAGA	ATCGAGGGCA	240
AGTAGTGAGA	GTGGTACTGT	TAAATCAGAG	TGATTATTGC	TAAGGTCTTT	GTAATTTGGG	300
GTTGTAGGTG	TTTTTTGTTT	TTGTTGTTTG	AGGGTCTGAA	TTTATTCGTT	ATATGATGTT	360
ATTGCCTGGA	ACTACCTTAT	CTGAGAAGCA	GTAGGCAATA	GAGTAGCGTA	TAAATGTTGG	420
TAAATTTTCT	CTTAAGGAAA	CAAATTATCC	TTACAAAATT	CCAAGTAAA	GAAATAAAGA	480
GAATGTATCT	TGGTTTTGTG	TGGAGAGAGG	GAAGTAGAAG	ATGGGGGATG	AAGAGAGAGA	540
GGAGGGTTAT	TTATTGGGCT	ATATATAGTG	TTGGTAGTAG	GAATCTTAAT	TCTTGTGTGT	600
AGTTTTGTTC	TTTTGTGTAT	AGTTATTGAT	TATTACTTTA	TTCCATGGGA	ATAATGAGTT	660
CCTATTATTT	CTGGAGGATA	TTTTGCCATT	TCGATGAGAC	ACACAGCCTC	TTCTTTGCTA	720
TGCAATATTA	CGAGATTACA	ACAGTTCTAA	CTCCCTGAAG	ACAAATACTT	CATGAGTCTC	780
ATTAGCTATC	TAAGCTATAG	GAAGAGCAGA	ATTTAATTCT	ACATGGAAAC	AGTAAGAAGC	840
TAGTATAATG	AAGAATTTTA	TTGATATCAC	TTGATTGAAA	TTTGTTCCTGA	CTCTTTAGAA	900
AAAGCAAGGG	TGAAATAAGA	TTTGTGATTC	TACAGTAGTA	ATGGGTAAGA	GGATAGGTCT	960
CAGGACAAAC	TGCCTAATGA	AACCCTAAAT	CTGTTATTTA	TTTATTTTCT	GATGACAGTG	1020
GGATAACTGA	CATTTACACA	TTAGCTTTCT	CATATGTAAA	AAAGAAATTT	TATTTTTTATT	1080
ATAGTCTGTC	AAGGAATATT	AAATATAAGG	TTTTGGAGCA	TGGTTGATAT	TTAGCAGATG	1140
TCTGTTTATT	CTTGATCAGT	ATAGAGTTGC	CACTTGGAAA	ATGCATCTTG	AAGATTACAT	1200
AACCAGACAA	AATTTGTTAG	TAACACTCAG	TGGTCTTAAG	ATGTTATAAG	TGACGGGCTA	1260
GTCGTGGTAA	TCAACTTGAT	ACCTTGACCC	TCAGGAGAAG	AGGGATTGTC	TCCATCGGAT	1320
GGGCTGTGA	GCATATCTGT	GGGGACGTTT	TTCTTGGACT	GCCTAGTTGA	TGGAAAAGGG	1380
CTTGGCTCAG	TGTCAGTGGT	CCTTCTTATG	GTGAGCAAGC	TGGGGGAAGC	GTTGCAGTAA	1440
GCAGTAGTCC	TTTGTGGTCT	CAGCTTCCTT	TTCTTCTCTC	TTCTTTCTTT	CTTTCTTTCT	1500
TTCTTTCTTT	CTTTCTTTCT	TTCTTCCTTC	CTTCCTTCCT	TTTCTCTCTT	TCTTTCTTTA	1560
GTTCCGTTTC	TTTGTTTATT	CGTTTCGTTT	TCGAGACAGG	GTTTTTCTGT	ATAGCCCTGG	1620
CTGTCCTGGA	ACTCACTTTG	TAGACCAGGC	TGTCCTCGAA	CTCAGAAATC	CGCCTGCCTC	1680
TGCCTCCCTG	TGAGTGCTGG	AATTAAAGGC	ATGCGCCACC	CCGCCCGGCT	TCTCAGCTTC	1740

CATTTCTGTT CAAGCTCTTG CCTTCAGCTC CTGCCTTGGC TTTCTGAGAC AAAGGCATAT	1800
AATCTGTAAG CCAAATCAAA CTTTCTTCT CAACTTGCTT TTGGCCAGTG TTTTATTACA	1860
GCGACTAAAG GCAAAC TAGA CTACTATGTA AATGGGAAGC ACTGTTAAAG TCAAGTAATA	1920
GCAAAAGATT ACATGGCCTG GATTTTTTGA GGTGCTTAC TTTCTCTGTG TACCCGGTTG	1980
TAAGTGTCTT TCCTACTTTT TTTATTAGCA TTTT TTTCC ATGTTTGTGCT TTGCACATAG	2040
AGAAGTTTGA AGCACTTTAT TTTGTAGGGT GTTTTGTATA ATCTGTCCAC CATCATTTTT	2100
ATTGTTTCT TATGTTTTTT CAAGATTCT TTGGGAGCCC TGGAAAC	2147

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Gly	Lys	Thr	Gln	Met
1				5

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CAGACGGTCA CACAGCTCTT GTGATAA

27

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ACCCACTCGT TTTAGTTTCT TGCTAC

26

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TAGAGAGAGA GAGAGAGCGA GACAG

25

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GTCGACCACG CGTGCCCTAT AG

22

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TCACTGCATC TGGAAGCAC

19

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CTAACTAAAT GGCGAGCATT GAG

23

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GAGCTCGAGG GTACCCATGG GTAGCAAGAA AC

32

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CGAGGAAGCA GT

12

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CGAGGAAGAA AA

12

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GGAGGCATTA AA

12

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GGTGGCGAAA TT

12

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CGTTACCCTA T

11

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

SGWGGMRRNA NA

12

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CGTT_CCCTA T

11

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CCTA_CCCTA A

11

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CGTTACCCTA T

11

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CGTTACCCAA T

11

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CGGTTGCCAT G

11

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CGTTACCCTA T

11

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GTAGGTCGAA

10

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CGAGGAAGCA

10

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

CGAGGAAGAA

10

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

CGTGGAGGCA

10

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

CGTGGTGGGA

10

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

CGAGGATGAC

10

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CGTGGTTGAT

10

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

CGAGGTCGCA

10

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CTAGGTAGCA

10

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GTAGGTGTTA

10

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CTAGGTAACA

10

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

CATGGTTGCC

10

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

CGWGGWNGMM

10

CLAIMS:

1. An isolated and purified nucleic acid that
 - a. encodes a protein comprising a sequence that is the sequence of a mammalian ATP-dependent homologous pairing protein or a protein substantially identical to the protein of SEQ ID NO: 1;
 - b. is labeled by random-primed hsREC2 cDNA in a Southern blot washed twice at 42°C for 20' in 2x SSC, 0.1% SDS followed by thrice at 50°C for 20'; and
 - c. contains a continuous coding sequence of at least 132 nucleotides of which greater than 100 are identical with a continuous 132 nucleotide sequence of SEQ ID NO: 2.
2. The nucleic acid of claim 1, which comprises a cDNA obtained from a species of mammal.
3. The nucleic acid of claim 1, which comprises a genomic DNA obtained from a species of mammal.
4. The nucleic acid of claim 3 which comprises the inserts of clones λ5A and λ1C, deposited as ATCC No. 97683 and No. 97682, respectively.
5. The nucleic acid of claim 3 that encodes a protein comprising the sequence of residues 2-350 of SEQ ID NO:1.
6. The nucleic acid of claim 1, in which the sequence of the pairing protein is greater than 90% identical to residues 4-347 of SEQ ID NO:1.
7. The nucleic acid of claim 6, which comprises the inserts of clones λ5D2a and λ7B1a, deposited as ATCC No. 97686 and No. 97684, respectively.
8. The nucleic acid of claim 6 that encodes a protein comprising a sequence that is

substantially identical to residues 4-347 of SEQ ID NO:1.

9. The nucleic acid of claim 5 having a sequence comprising the sequence of bp 74 to 1120 of SEQ ID NO:2.

10. The nucleic acid of claim 9, which further comprises a promoter.

11. The nucleic acid of claim 10, which is pCHsREC2 deposited as ATCC No. 97685.

12. A nucleic acid having a sequence which comprises a fragment of at least 20 nucleotides of SEQ ID NO:2 or SEQ ID NO:4 or a complement thereof and a label.

13. A kit comprising:

- a. a 5'-nucleic acid fragment having a sequence which comprises a 5'-sequence of at least 12 nucleotides of SEQ ID NO:2; and
 - b. a 3'-nucleic acid fragment having a sequence which comprises a 3'-sequence of at least 12 nucleotides of the complement of SEQ ID NO: 2;
- wherein

the 3'-sequence is complementary to a portion of SEQ ID NO:2 that is 3' to the 5'-sequence.

14. A composition which comprises an ATPase, which composition is substantially free of other normally intracellular mammalian proteins, in which the sequence of the ATPase comprises a sequence that is substantially identical to a continuous sequence, at least 120 amino acids in length, of a mammalian ATP-dependent homologous pairing protein, and in which the ATPase is an ATP-dependent homologous pairing protein.

15. The composition of claim 14, in which the ATPase is an mREC2.

16. The composition of claim 14, which comprises an ATPase having a sequence which comprises at least 115 amino acids of SEQ ID NO:1 or which is substantially

identical thereto.

17. A composition which comprises an ATPase, which composition is substantially free of other normally intracellular mammalian proteins, in which the sequence of the ATPase comprises a sequence that is substantially identical to the sequence of residues 80-200 of SEQ ID NO:1, and in which the ATPase is an ATP-dependent homologous pairing protein.

18. The composition of claim 17, in which the ATPase is an mREC2. .

19. The composition of claim 18, in which the ATPase is substantially identical to hsREC2.

20. The composition of claim 19, wherein the sequence of the ATPase comprises amino acids 2-350 of SEQ ID NO:1.

21. A method of classifying a sample of human tissue, which comprises:

- a. quantifying the copies of *hsREC2* per diploid genome of a sample tissue;
and
- b. comparing the quantity of *hsREC2* per diploid genome of the sample tissue with the quantity of *hsREC2* per diploid genome of a standard tissue.

22. The method of claim 21 wherein the comparison is performed by measuring the lengths of microsatellite DNA at marker D14S258 and comparing the sizes present in the sample tissue and the sizes present the standard tissue, provided the standard tissue and the sample tissue are from the same subject.

23. A method of classifying a sample of human tissue, which comprises comparing a *hsREC2* gene of a sample tissue with a *hsREC2* gene of a standard tissue.

24. The method of claim 23 wherein the comparison is performed by determining the

presence or absence of a single stranded conformational polymorphism between the *hsREC2* genes of the sample and of a standard tissue.

25. The method of claim 23 wherein the comparison is performed by obtaining the sequence of a fragment of the *hsREC2* of the sample tissue and comparing the obtained sequence with the sequence of SEQ ID NO:2 or or a complement thereof.

26. A transgenic mouse having at most one copy of *muREC2* per diploid genome that encodes a *muREC2* protein.

27. The transgenic mouse of claim 26 having no gene that encodes a *muREC2* protein.

28. A transgenic animal comprising an *mREC2* gene operably linked to a heterologous promoter such that the *mREC2* gene is expressed.

29. The transgenic animal of claim 28, in which the promoter is a tissue specific promoter or an inducible promoter.

30. An embryonic stem cell line comprising an *mREC2* gene operably linked to a heterologous promoter such that the *mREC2* gene is expressed.

31. The embryonic stem cell line of claim 30, in which the promoter is a tissue specific promoter or an inducible promoter.

32. An antibody or fragment thereof which binds a protein having a sequence of SEQ ID NO:1 and binds to no other human protein.

33. A method of making a specific genetic alteration in a mammalian cell which comprises:

- a. increasing the level of *mREC2* in the cell; and
- b. introducing into the cell a mutation-containing nucleic acid having

a region of homology with the genome of the cell,
such that the nucleic acid and the genome of the cell homologously recombine causing
the mutation in the genome.

34. The method of claim 33 wherein step (a) comprises transporting an exogenous
nucleic acid that encodes a mREC2 into the cell.

35. The method of claim 33 wherein step (a) comprises transporting exogenous
mREC2 protein into the cell.

36. The method of claim 33 wherein the mutation-containing nucleic acid is a CMutV.

37. A composition comprising an isolated and purified mammalian REC2 promoter.

38. The composition of claim 37, wherein the REC2 promoter is a radiation
induceable promoter.

39. The composition of claim 37, wherein the REC2 promoter is operably linked to an
enhancer.

40. The composition of claim 37, wherein the REC2 promoter is operably linked to a
gene encoding a protein other than a mammalian Rec2 protein.

41. The composition of claim 37, wherein the REC2 promoter is operably linked to a
gene encoding a Herpes Virus thymidine kinase gene.

42. The composition of claim 37, which further comprises a bacterial cloning plasmid
that contains the REC2 promoter.

43. A composition comprising a mammalian, radiation induceable REC2 promoter
operably linked to a strong enhancer.

44. The composition of claim 43, in which the composition further comprises a mammalian cell.
45. The composition of claim 43, in which the REC2 promoter is a hsREC2 promoter.
46. The composition of claim 43, wherein the enhancer is selected from the group consisting of the SV40 enhancer, the Hepatitis B Virus enhancer, the Cytomegalovirus enhancer and the α -fetoprotein enhancer,
47. The composition of claim 43, wherein the composition is a mammalian cell.

ABSTRACT

The invention concerns mammalian recombinase genes (REC2) and their promoters. Over expression of REC2 in a cell is found to facilitate homologous recombination, particularly homologous recombination using a DNA/RNA chimeric oligonucleotide and to sensitize a cell to the apoptotic effects of irradiation. The REC2 promoter, in combination with a strong enhancer, e.g., a SV40 enhancer, was found to be a strong promoter following irradiation of the cells. A radiation induceable promoter can be used to sensitize a cell to radiation treatment by operably linking the radiation induceable promoter to a gene whose expression converts a prodrug to a drug such as a herpes thymidien kinase gene.

Pending Application 60/135,139

Filed on 05/21/1999

Inventor: Bartlett

Our file: 7991-050-888

ATTORNEY DOCKET NUMBER: 7991-086-99
SERIAL NUMBER: 09/685,403
REFERENCE: **BG**

POINT MUTATION REVERSION IN THE DYSTOPHIN GENE

5 TECHNICAL FIELD

This invention relates to the field of muscular dystrophy and methods for its treatment in humans. The present invention also concerns the canine model of Duchenne muscular dystrophy in golden retrievers (GRMD). Another aspect concerns chimeric RNA/DNA
10 capable of inducing reversion of the point mutation causing (GRMD).

BACKGROUND

Chimeric Mutation

15 The inclusion of a publication or patent application in this section is not to be understood as an admission that the publication or application occurred prior to the present invention or resulted from the conception of a person other than the inventors.

An oligonucleotide having complementary deoxyribonucleotides and ribonucleotides and containing a sequence homologous to a fragment of the bacteriophage M13mp19, was
20 described in Kmiec, E.B., et al., November 1994, Mol. and Cell. Biol. 14, 7163-7172. The oligonucleotide had a single contiguous segment of ribonucleotides. Kmiec et al. showed that the oligonucleotide was a substrate for the REC2 homologous pairing enzyme from *Ustilago maydis*.

Patent publication WO 95/15972, published June 15, 1995, and corresponding U.S.
25 patent application Serial No. 08/353,657, filed December 9, 1994, by E.B. Kmiec, now U.S. Patent No. 5,565,350 (the '350 patent) described duplex CMV for the introduction of genetic changes in eukaryotic cells. Examples in a *Ustilago maydis* gene and in the murine ras gene were reported. The latter example was designed to introduce a transforming mutation into the ras gene so that the successful mutation of the ras gene in NIH 3T3 cells would cause the
30 growth of a colony of cells ("transformation"). The '350 patent reported that the maximum rate of transformation of NIH 3T3 was less than 0.1 %, i.e., about 100 transformants per 10⁶

cells exposed to the ras duplex CMV. In the *Ustilago maydis* system the rate of transformants was about 600 per 10⁶. A chimeric vector designed to introduce a mutation into a human bcl-2 gene was described in Kmiec, E.B., February 1996, Seminars in Oncology 23, 188.

- 5 A duplex CMV designed to repair the mutation in codon 12 of K-ras was described in Kmiec, E.B., December 1995, Advanced Drug Delivery Reviews 17, 333-40. The duplex CMV was tested in Capan 2, a cell line derived from a human pancreatic adenocarcinoma, using LIPOFECTIN™ to introduce the duplex CMV into the Capan 2 cells. Twenty four hours after the duplex CMV were introduced, the cells were harvested and genomic DNA
10 was extracted; a fragment containing codon 12 of K-ras was amplified by PCR and the rate of conversion estimated by hybridization with allele specific probes. The rate of repair was reported to be approximately 18%.

- A duplex CMV designed to repair a mutation in the gene encoding liver/bone/kidney type alkaline phosphatase was reported in Yoon, K., et al., March 1996, Proc. Natl. Acad.
15 Sci. 93, 2071. The alkaline phosphatase gene was transiently introduced into CHO cells by a plasmid. Six hours later the duplex CMV was introduced. The plasmid was recovered at 24 hours after introduction of the duplex CMV and analyzed. The results showed that approximately 30 to 38% of the alkaline phosphatase genes were repaired by the duplex CMV.

- 20 United States Patent Application Serial No. 08/640,517, filed May 1, 1996, by E.B. Kmiec, A. Cole-Strauss and K. Yoon, published as WO 97/41141, November 6, 1997, and the publication Cole-Strauss, A., et al., September 1996, Science 273, 1386 disclose duplex CMV that are used in the treatment of genetic diseases of hematopoietic cells, e.g., Sickle Cell Disease, Thalassemia and Gaucher Disease. United States Patent Application Serial
25 No. 08/664,487, filed June 17, 1996, by E.B. Kmiec, published as WO 97/4871, December 24, 1997, describes duplex CMV having non-natural nucleotides for use in specific, site-directed mutagenesis. The duplex CMV described in the applications and publications of Kmiec and his colleagues contain a central segment of DNA:DNA homoduplex and flanking segments of RNA:DNA heteroduplex or 2'-OMe-RNA:DNA heteroduplex. Kren et al.,

1997, Hepatology 25, 1462-1468, reports the successful use of a CMV in non-replicating primary hepatocytes.

United States provisional applications Serial No. 60/054,837, August 5, 1997, Serial No. 09/108,006, June 30, 1998, and No. 60/064996, filed November 5, 1997, concern the
5 use of CMV in non-replicating cells and compositions of CMV and macromolecular carriers, including macromolecular carriers that have ligands for clathrin-coated pit receptors.

The Introduction of DNA into Muscle Cells

10 There are several references that report the introduction and expression of plasmid DNA encoding the dystrophin protein into skeletal muscle. Acsadi G. et al., 1991, Nature. 352(6338):815-8; Danko, I. et al., 1993, Human Molecular Genetics. 2(12):2055-61; Bartlett, R.J., et al., 1996, Cell Transplantation 5:411-419; Wells D. J. et al., 1993, FEBS Letters. 332(1-2):179-82; Fritz J.D. et al., 1995, Pediatric Research. 37(6):693-700; Wolff
15 JA. et al., 1992, Hum Mol Genet. 1:363-9; Inui, K. et al., 1996 [Review] Brain & Development. 18:357-61. A general method of introducing DNA into a muscle cell for the purpose of inducing an immune response in a host is disclosed in U.S. Patent No. 5,589,466 and No. 5,580,859, both to Felgner et al. The expression of an exogenous dystrophin gene is an example in these patents.

20 Experiments directed at determining a ligand that can be used to introduce large DNA fragments into the myofibers of DMD patients are reported by Feero, W.G., et al., 1997, Gene Therapy 4, 664-674. The use of liposomes to deliver DNA to myofibers for expression without the use of a targeting ligand is described in Smyth-Templeton, N., et al., 1997, Nature Biotechnology 15, 647.

25

The molecular Biology of Muscular Dystrophy

The muscular dystrophies comprise a genetically and clinically diverse set of diseases characterized by abnormalities of the skeletal muscle. Reviewed Straub, V., et al., 1997, Cur. Op. Neurology 10, 168-175. The muscular dystrophies can be classified by the

30

mode of inheritance, autosomal dominant, autosomal recessive and X-linked, and each type further divided according to the chromosomal locus and even the effected gene, if known.

The most common muscular dystrophy is X-linked with the dystrophin gene effected. The dystrophin gene occupies 2,300 kb or about 1.5 % of the X-chromosome. Its
5 mature transcript is 14 Kb and encodes a protein of 3685 amino acids having a molecular weight of 427 kd. The gene contains 79 exons. The dystrophin gene is extraordinarily large; it is about half the size of an *E. coli* genome. There is no clear explanation for its size. Reviewed, Worton, R.G., & Brooks, M.H., 1995, *The Metabolic and Molecular Basis of Inherited Disease 7th Ed.* Chapter 140 (McGraw Hill, New York).

10 The dystrophin protein contains an N-terminal binding region, that binds to intracellular filamentous actin (which is not the actin of the contractile apparatus), a C-terminal binding domain that binds to a transmembranous glycoprotein complex which in turn binds to laminin, and a connective region. Under physiologic conditions dystrophin exists as a homodimer, which connects the actin filaments with the glycoprotein complex as
15 well as linking each.

Although there are multiple mutations of dystrophin that result in muscular dystrophy, the mutations can be classified into types. The milder form, termed Becker Muscular Dystrophy (BMD), is associated with genomic deletions or mRNA processing errors that do not alter the reading frame of the mature mRNA and, hence result in a mutant
20 protein that contains intact N terminal and C terminal binding domains. In the more severe form, termed Duchenne Muscular Dystrophy (DMD), the dystrophin protein lacks a C-terminal binding domain and is usually unstable. DMD typically results from point mutations that introduce in frame termination codons or from insertion or deletion mutations that result in a frame-shift. Koenig, M., 1989, *Am. J. Hum. Gen.* 45, 498; Prior, T.W., et al.,
25 1995, *Human Mutation* 5, 263; Koenig, M., et al., 1987, *Cell* 50, 509; Baumbach, L., et al., 1989, *Neurology* 39, 465.

The relationship between the pathophysiology of DMD and BMD and the physiologic function of dystrophin is complex. Dystrophin is not required to transmit the force of the contractile apparatus to the tendonous connections of the muscle. Rather the
30 defective muscles undergo an ongoing series of focal necrosis of the myofibers, which

ultimately exceed the repair capacity of the muscle. The end stage disease is characterized by fibrosis between myofibers, atrophy and weakness.

Dystrophin Replacement Gene Therapy

5

Several groups have attempted to treat DMD by introducing genes encoding dystrophin into the myofibers of affected individuals. A variety of methods have been employed. The methods can be classified into three groups: *in situ* replacement gene therapy; *ex vivo* replacement gene therapy using autologous myoblasts, which are then
10 reimplanted; and allogenic transplantation of wild-type myoblasts.

Examples of the first type include the above-noted transfections of differentiated myofibers using DNA and non-biologic carriers. This form of therapy has been of limited value to date because of the low efficiency of transfection. The use of adenovirus based vectors to increase efficiency has been reported. Vincent, N. et al. 1993, Nature Genetics.
15 5:130-4; Ragot, T. et al., 1994, Gene Therapy. 1 Suppl 1:S53-4; Acsadi et al., 1996, Human Gene Therapy. 7:129-40; Deconinck, N. et al., 1996, Proc. Natl. Acad. Sci. 93:3570-4; Clemens, P.R., et al., Gene Therapy. 3:965-72; Haecker, S.E. et al., 1996, Human Gene Therapy. 7:1907-14; Chen, H.H., et al., 1997, Proc. Natl. Acad. Sci. 94:1645-50; Yang Y., et al., 1995, Journal of Virology. 69:2004-15; Haecker, S.E., et al., 1996,
20 Human Gene Therapy. 7:1907-14. Although efficiencies as high as 50% have been reported in experimental animal systems, Ragot, T., et al., 1993, Nature 361:647, adenovirus-based therapies have likewise been of limited value to date because the expression of dystrophin has been transient and there is an immune response to the adenovirus vector that limits the possibilities of repeated therapy. Although gene therapy has not proved to be a practical
25 clinical modality, it has been useful to demonstrate that the expression of a wild-type dystrophin in an DMD model system results in amelioration of the disease. Danko, I., et al., 1993, Human Mol. Genetics 2, 2055-61.

Techniques for the culture of myoblasts from normal individuals have been reported. U.S. Patent No. 5,538,722 to Blau & Hughes. The transfer of dystrophin into cultured
30 myoblasts has been reported, Dunckley, M.G., et al., 1992, FEBS Letters 296:128-34,

however, this approach has not been pursued because a secondary effect of DMD is a decline in the numbers of myoblasts that can be recovered in culture, Webster, C. & Blau H.M., 1990, Somatic Cell & Mol. Genetics, 16:557-65. Allegedly, successful engraftment of allogenic cultured myoblasts have been reported. U.S. Patent No. 5,130,141 to Law; 5 Law, P.K., et al., 1992, Cell. Transplantation 1,235.

However, other studies, conducted under controlled conditions, have failed to confirm the clinical benefit of allogenic myoblast grafts. Gussoni, E., et al., 1992, Nature 356, 435; Karpati, G. et. al., 1992, Ann. Neurol. 34:8. There is consequently a need for a therapy that effects the long-term expression of dystrophin in muscle fibers of DMD 10 subjects. Ideally the therapy should be deliverable to all tissues through intravenous or intraarterial injection.

A further limitation of both myoblast engraftment and non-viral-based gene therapy is a requirement for local delivery, such that multiple injections are required to treat even a single large muscle. Reports to the contrary with regard to myoblast engraftment, 15 notwithstanding, *e.g.*, Hughes, S.M., et al., 1990, Nature 345, 350; Neumeyer A.M., et al., 1992, Neurology 42, 225, more recent studies have not confirmed that transvascular engraftment into muscle fibers occurs to any practical extent.

Two well characterized animal models exist for Duchenne muscular dystrophy, the *mdx* mouse [Bulfield, G. et al. Proc. Natl. Acad. Sci. USA 81:1189-1192, 20 1984 and Sicinski, P. et al. Darlison, M.G., and Barnard, P.J. Science 244:1578-1579, 1989] and the golden retriever dog [Kornegay, J.N. et al. Muscle and Nerve 11:1056-1064, 1988 and Sharp, et al. Genomics 13:115-121, 1992]. In both cases, a point mutation has been identified as causal, with the mouse having a nonsense mutation in exon 23, and the dog having a splice acceptor site mutation in intron 6 causing a frame-shift due to complete 25 deletion of exon 7 from the mature canine dystrophin mRNA [Wilton, S.D. et al. N.G. Muscle and Nerve 20:728-734, 1997; Wilton, S.D. et al. Neuromuscular Disorders: 7:329-335, 1997; and Schatzberg, S.J. et al. Muscle and Nerve 21:91-998, 1998.] Alternate splicing mechanisms, which restore the dystrophin reading via removal of mutation containing out-of-frame exons, have been suggested to play a causal role for the presence of 30 dystrophin positive staining "revertant fibers" in both models, although no evidence of true

reversion of these point mutations at the genomic level have been reported. A considerable amount of effort has gone into the study of gene therapy in the *mdx* model using direct DNA injection [Acsadi G. et al. *Nature*. 352(6338):815-8, 1991.] viral vectors [Danko, In. et al. *Human Molecular Genetics* 2:2055-61, 1993 and Wells, D.J. *FEBS Letters*. 332:179-82, 1992.] and myoblast transplantation [Fritz, J.D. et al. *Pediatric Research* 37:693-700, 1995 and Inui, K., et al. *Brain & Development*. 18:357-61, 1996.] with modest levels of short-term success due to limitations of transfection targeting and efficiency, and either acute or chronic immune responses directed against cells which express the therapeutic gene product [Kinoshita, In. et al. *Acta Neuropathologica*. 91:489-93, 1996; Kinoshita, In. et al. *Neuromuscular Disorders* 6:187-93, 1996; Yang, Y. et al. *Journal of Virology* 70:7209-12, 1996; Yang, Y. et al. *Gene Therapy*. 3:137-44, 1996; and Worgall, S., et al. *Human Gene Therapy*. 8:37-44, 1997.]. Recent studies have suggested that myoblast transplantation therapy of Duchenne muscular dystrophy is also ineffective [Partridge, T. et al. *Nature Medicine* 4:1208-1209, 1998 and Mendell, J.R. et al. *New England J. Med.* 333:832-838, 1995.]. Long-term correction of dystrophin deficiency requires a more permanent gene therapy which will provide stable expression of dystrophin absent of problems associated with vector or cell delivery.

Recently, [Cole-Strauss et al. *Science* 273:1386-9, 1996.] a novel chimeric RNA and DNA oligonucleotide (chimeraplast) was used to target reversion of the sickle hemoglobin allele in transformed Sickle lymphoblast cell-line. This technique, termed chimeraplasty, is believed to rely on the sequence homology designed into the chimeraplast bracketing the site of the chromosomal mutation to direct the host cell nuclear mismatch repair proteins to revert the chromosomal sequence to that designated within the chimeraplast [Ye, S. et al. *Mol Med Today*: 4:431-7, 1998.]. In the sickle cell study, this resulted in the reversion to wild-type of 20% of sickle chromosomes.

A critical issue in the field of gene therapy is reliable and safe introduction of genes and oligonucleotide fragments into the subject's cells. Transport of the highly charged oligonucleotide fragments through the cell membrane has proved challenging, and current protocols are very limited and generally laborious.

We investigated whether: 1) a chimeraplast designed to revert the GRMD splice-site mutation (Figure 1) could restore inclusion of exon 7 into the GRMD dystrophin mRNA in skeletal muscle of live animals; and 2) a method and composition effective for introducing the chimeraplast into the affected cell.

5

SUMMARY OF THE INVENTION

Multiple lines of evidence confirm that direct *in vivo* injection into dystrophic skeletal muscle of an appropriately designed and synthesized chimeric RNA/DNA oligonucleotide (GRMD chimeraplast) is capable of inducing reversion of the point mutation causing GRMD and in humans having Duchenne and Becker type muscular dystrophy. We have also surprisingly found that FuGENE™ 6 is an effective carrier vehicle for sustained inclusion in nascent dystrophin mRNA of epitope expression of exon 7. A composition comprising the chimeraplast packaged in FuGENE™ 6 was successfully introduced into the cell and produced dystrophin protein containing exon 7 epitopes. The invention further encompasses the use of alternative lipid carriers that are equivalent to FuGene™ 6, now known or to be developed.

Reversion of the GRMD point mutation, as measured at the mRNA, protein, and genomic DNA levels, was found up to 11 months after a single treatment with chimeraplast. To facilitate the study of the GRMD model, an exon seven-specific antibody against the portion of the protein deleted by the GRMD mutation provided a unique reagent for discriminating patterns of dystrophin protein expression resulting from chimeraplasty to that produced by alternate processing of the mRNA. The critical importance of exon-specific antibodies for unequivocal identification of wild-type dystrophin in muscle fibers has been demonstrated in human myoblast therapy trials. At 11 months post-injection, detectable quantities of normal sized dystrophin were localized in multiple regions within the treated cranial tibialis muscle using the MANEX7B antibody. These results were revealed by both western blot and immunohistochemical analyses using the MANEX7B antibody. We estimate that the level of reversion approaches, but does not exceed, 1% in our studies based

on comparative levels of RT/PCR product from the exon 7 deleted mRNA produced by the GRMD allele in the 9 week biopsy sample. To clarify these analyses, RT/PCR primers were specifically selected to discriminate the mutant mRNA and reverted mRNA species from alternately spliced products. Precise quantitative estimates of the level of reversion have
5 proven difficult due to the inherent AT-rich nature of the intron portion of this splice junction, which renders a quantitative method such as allele-specific primer discrimination problematic at best. Thus, we have been limited to qualitative differences rather than quantitative differences between the mRNA/genomic results from the tissues treated with the two chimeraplasts used in these experiments versus untreated tissue from the same animal.
10 It is of interest to note that RT/PCR of RNA extracted from the necropsy samples from the right limb treated with the chimera without FuGENE™ 6 failed to produce any detectable exon 7 containing dystrophin mRNA. This is in contrast to the localization seen in both frozen sections taken from the small biopsy sample taken at 6 weeks for the in situ RT/PCR as well as the immunohistochemistry of the 6 weeks sample. Based on this contrast, we
15 believe that the initial reversion frequency for the two different limbs favored the delivery using the FuGENE™ 6 carrier vehicle for sustained inclusion in nascent dystrophin mRNA of epitope expression of exon 7.

Since the gene therapy vehicle used in these studies, a chimeric oligonucleotide, actually modifies the mutant gene with all of the native control elements for dystrophin
20 expression, production of dystrophin from a threshold level of reverted genes would be predicted to permit normalization of dystrophin expression patterns in the skeletal muscle. Expanded studies with multiple animals would also permit force generation analyses to correlate potential strength improvement produced from expression of normalizes dystrophin. Moreover, as the resulting dystrophin gene expression patterns reported here are
25 subclinical, exploration of methods to improve frequency of reversions are under study. These improvements would include: 1) using higher concentrations of chimeraplast delivered either as a single bolus, 2) serial treatments, 3) extended delivery via an implantable osmotic pump and 4) alternate methods of physical introduction such as electroporation, and use of a carrier molecule such as a modified polyethyleimine. Based
30 on a previous report in liver using a chimera to mutate the factor IX gene in rats, higher

levels of gene modification were achievable by improving chimeraplast delivery. The putative clinically relevant threshold of dystrophin expression to prevent the dystrophic phenotype has been suggested to be 20% of normal levels. Thus, strategies which produce higher levels of reversion may be useful since chimeraplasts have little inherent capacity for inducing an immune response. As reported previously for liver, serial administration of chimeraplasts in dystrophic muscle may have additive effects and may result in achievement of clinically relevant levels of gene modification which would be measurable by force-generation in this animal model for Duchenne muscular dystrophy.

Furthermore, we believe the GRMD model should also be useful for studying the potential for using chimeraplasts for restoration of reading frame caused by deletions. The fact that exon 7 is missing from the dystrophin mRNA in dogs with this mutation actually simulates an exon seven genomic deletion. Thus, a chimeraplast designed to restore reading frame by modifying the coding sequence beginning in exon 8 to match the reading frame from exon 6 would be predicted to produce a protein that would be Becker-like and may have sufficient function to normalize the muscle in this model. These studies are in the early stages of implementation.

DESCRIPTION OF DRAWINGS

These and other features, aspects, and advantages of the present invention will become better understood with regard to the following description, appended claims, and accompanying drawings where:

Fig. 1 shows the chimeraplast used for GRMD mutation reversion;

Fig. 2 shows the RT/PCR of total RNA from canine skeletal muscle;

Fig. 3 shows the in situ RT/PCR of canine muscle;

Fig. 4 shows the epitope mapping of dystrophin antibodies and western blotting of canine skeletal muscle;

Fig. 5 shows the immunofluorescence localization of dystrophin in canine muscle; and

Fig. 6 shows the genomic PCR analysis of mutation reversion.

EXAMPLES

The following examples are provided for illustrative purposes only and are not to be construed as limiting the invention's scope in any manner.

5 Example 1: Chimeraplast Preparation and Injection into Dystrophic Skeletal Muscle

To determine if FuGENE™ 6 mediated chimeraplasty could be used to revert the point mutation responsible for GRMD, a six week old affected male was selected for study. After surgical exposure of the sartorial compartment, 5 mls of injectate containing reverting chimeraplast (200µg from BioSource), 200 µg of calf thymus histone H1 (Sigma) and
10 packaged in FuGENE™ 6 (Roche Diagnostics Corporation), was delivered in 50 separate 100 µl injections. A biopsy of control untreated triceps muscle was removed for RNA, western, and immunohistochemical studies prior to injection. Times for follow-up biopsies and additional injection of chimera are diagrammed in Figure 1B.

We found that FuGENE™ 6 is an effective carrier vehicle for sustained inclusion in
15 nascent dystrophin mRNA of epitope expression of exon 7, while the same composition not packaged in FuGENE™ 6 was ineffective and produced no dystrophin protein containing exon 7 epitopes. FuGENE™ 6 (<http://biochem.roche.com/techserv/fugene.htm>) is commercially available from Roche Diagnostics Corporation (Indianapolis, IN.) FuGENE™ 6 is a proprietary blend of lipids and other components supplied in 80% ethanol,
20 sterile filtered, and packaged in polypropylene tubes.

Example 2: RT/PCR Analysis of Dystrophin mRNA in Treated Skeletal Muscle

To investigate whether therapy with reverting chimeraplast produced a change in the pattern of dystrophin gene expression in GRMD muscle, total RNA was isolated from frozen
25 sections of biopsy and necropsy material taken at various timepoints after treatment. RT/PCR analysis was performed using primers which bracket exon 7 in the canine dystrophin mRNA (Figure 2A). While suggestive levels of normal sized dystrophin RT/PCR product containing exon 7 were seen in the 2 week sample, the results from the 9 week sample demonstrated that at least as much product from normal sized mRNA was
30 present in the biopsy as the mutant mRNA. (Figure 2B) Confirmation of the presence of

exon 7 in the PCR product was by sequencing and re-PCR with a exon 7 specific 3' primer (Figure 2A) and the original 5' primer (data not shown). Moreover, analysis of the necropsy sample from the left limb (FuGENE™ 6 treated sample) taken at 11 months reveals that the predominant RT/PCR product was of normal size. Since the level of mutant mRNA is <1% of normal in muscle of GRMD dogs and not visible on a Northern blot, (data not shown) we surmise that the reversion frequency in the studies of this animal produced a similar modest level of normalized dystrophin mRNA.

Example 3: In situ RT/PCR of Treated Skeletal Muscle

To determine what the pattern of distribution of reversion was in the injected muscle, we performed *in situ* RT/PCR on frozen sections from normal, GRMD muscle and the 6 week injected sample from the right leg. Examination of negative control sections from GRMD triceps muscle obtained via biopsy prior to injection revealed complete absence of exon 7 across the entire section (Figure 3A). Examination of positive control sections from normal canine muscle showed localization of exon 7 across the entire section (Figure 3B). Experimental sections from injected GRMD muscle had modest localization of exon 7 across the entire section particularly near to fluorescent microspheres indicating proximity to sites of injection (Figures 3C and 3D). At high magnification, the injected samples show discrete localization of exon 7-containing dystrophin mRNA at the periphery of fibers where one would expect the myonuclei to be located. (Figure 3 E-H) These results suggest that modest reversion occurred in multiple nuclei proximal to the sites of injection.

Example 4: Preparation of Exon Seven-Specific Monoclonal Antibodies

Initial Western blotting and histochemical analysis of the 2 and 9 week samples obtained from tissue from the left limb as well as the 6 week sample taken from the right limb using a commercially available carboxy-terminal dystrophin antibody (Novacastra) suggested no apparent increase in dystrophin protein and modest evidence of dystrophin positive fibers located in the region of the injection site marked by fluorescent microspheres. However, the levels were no higher than background reversion when compared to uninjected sample from the triceps muscle taken from the same animal prior to therapy. (data not

shown) To increase specificity in the immunological analyses, an exon 7-specific antibody was used. Fourteen mAbs raised against a fragment of dystrophin encoded by exons 4-16 were mapped by Western blotting with sub-fragments produced by PCR (See Figure 4A). The exon 7 mAbs, for example, recognized an exon 7-16 fragment, but did not recognize
5 exon 8-16 or any smaller fragment. This shows that exon 7 is essential for binding, and we may be confident that the exon 7 mAbs will not recognize "revertant" dystrophins lacking exon 7. The MANEX7B mAb was selected for further studies.

10 Example 5: Western Blotting of Treated Skeletal Muscle

To investigate whether increases in RT/PCR product containing exon 7 correlated with restoration of normal dystrophin, western blot analyses were performed using the MANEX7B mAb. When samples taken at necropsy were studied using this antibody, restoration of normal sized dystrophin protein containing exon 7 epitope was observed.
15 (Figure 4B) This is indicative that the treatment with chimera produced a modest level of gene reversion detectable at 11 months post injection. While both the left cranial tibialis (CT) muscle, in particular, and the long-digital extensor (LDE) muscle, to a lesser extent, revealed the expected high molecular weight band co-migrating with the normal muscle sample, no significant high molecular weight of dystrophin protein containing exon 7
20 epitopes was found in the right limb at necropsy (Figure 4B). As expected, no high molecular weight protein was found in the untreated GRMD muscle samples (Figure 4B). Due to limitation of sample size no 2, 6 or 9 week samples could be included in these studies. However, expression of the normal sized dystrophin protein containing the exon 7 epitope found 11 months after treatment with chimeraplast, provided confirming evidence
25 that modest reversion of the GRMD allele had occurred in the left leg.

Example 6: Fluorescent Immunohistochemistry of Treated Skeletal Muscle

To determine the pattern of dystrophin distribution in the treated skeletal muscle, frozen sections taken at necropsy were treated with MANEX7B primary and an FITC-
30 labeled secondary antibody raised in goats (Sigma). Upon scanning untreated triceps muscle

for any localization, none was found confirming the specificity of the MANEX7B antibody (Figure 5A and D). In contrast, peripheral staining of a small percentage of fibers was observed in the sections taken from both the right (Figure 5B) and left (Figure 5E) CT muscles, while the positive control muscles demonstrated a pattern of normal muscle staining of wild-type dystrophin (Figure 5C and F). As each injected muscle received numerous injections, positive fibers were found in clusters within the proximity of the sites of injection and usually were no more than 2-3 mm from an injection site. Due to limiting sample mass, biopsy samples from the 2 and 9 week were not tested. It is of interest to note that no exon 7 epitope was found in the right CT muscle at necropsy (data not shown).

10 However, the localization of the exon 7 epitope to the periphery of muscle fibers 11 months after treatment of the left CT muscle further confirms that reversion of the GRMD allele occurred via chimera treatment. The difference between the two treatments was the presence of FuGENE™ 6 in the left limb used as a carrier molecule. Based on similar results from parallel studies reported previously in the mdx mouse, we surmise that the

15 chimera was more readily transferred to the myonuclei using the FuGENE™ 6 carrier, and thus was able to sustain higher levels of reversion long-term.

Example 7: Genomic PCR of Exon Seven from Treated Skeletal Muscle

To confirm that the chimeraplastic process had actually reverted the mutation,

20 genomic DNA was isolated from additional serial frozen sections taken from the indicated muscles. Initial PCR primers and conditions were as previously described for the diagnosis of carriers of the GRMD allele. The GRMD mutation produces a novel Sau96 recognition site and digestion of the 310 bp genomic PCR product is diagnostic of the mutant allele. To enhance the likelihood that a reverted allele could be detected via amplification by PCR,

25 reactions were stopped after cycle 10, submitted to digestion with Sa96, extracted with phenol/chloroform and precipitated from ethanol. This precipitate was used as template for an additional 25 cycles of PCR, and the products cloned into a commercial PCR cloning vector (pCR-1, Invitrogen). Examination of 50 clones from the left CT muscle produced 3 that demonstrated a pattern of digestion with Sau96 indistinguishable from that obtained

30 from a normal dog muscle sample. (Figure 6A) These three clones were sequenced and

each contained the reverted sequence containing the functional splice acceptor site.

(Figure 6)

Example 8: Animal

- 5 An affected male (6 weeks of age) from a litter born at the University of Missouri colony was selected for study. All animals are maintained in the University of Missouri Vivarium according to ACUC and NIH guidelines for use of animals in research. At 13 months of age, disease symptoms warranted euthanizing the animal. All surgical biopsy and necropsy samples from all treated sartorial compartment muscles as well as the left
- 10 triceps were collected, wrapped in aluminum foil, and snap-frozen in liquid nitrogen.

Example 9: Direct Injection of oligonucleotide

- Direct injection of oligonucleotides into skeletal muscle was as reported previously for plasmid DNA. The sequence of the chimeraplast is shown in Figure 1. In the dog, the
- 15 GRMD mutation is predicted to produce a mismatch with the GRMD mutation sequence, and thus be reverted.

Example 10: RT/PCR of skeletal muscle RNA

- To isolate RNA, serial frozen sections of 20 μ M thickness (10-20) from the same
- 20 segments of muscle used for western blotting and immunohistochemistry were made and stored at -80°C in separate RNase-free tubes. Total RNA was isolated according to the protocol provided by the supplier (Qiagen). The method for RT/PCR was essentially as reported in the original paper defining the GRMD mutation.⁴

- 25 In situ RT/PCR of Frozen muscle sections. Frozen sections of muscle from normal, GRMD and GRMD injected muscle were prepared on Superfrost slides using a Lecia 3000 cryomicrotome. After overnight fixation in 10% buffered formalin, slides were rinsed 2 x in fresh PBs, then digested for 17 minutes in 5 μ g/ml pepsin, slides were then rinsed in 2 x changes of fresh PBS, and then treated with RNase free-DNAase to remove nuclear DNA,
- 30 and finally rinsed in 4 changes of fresh PBS. The slides were covered using in situ

chambers (RPI, Sci.) and a master mix containing the single enzyme TthI for performing both RT and PCR in a single tube was used, both 3' primers within exon 7 (M23 and 459), and primer 354 were used as primers in the presence of dNTP's and biotinylated dATP. After RT/PCR, slides were rinsed 2 x with fresh PBS, then treated at room temperature with streptavidin alkaline phosphatase derivative to bind the in situ biotinylated PCR product. Then the slides were again rinsed with 3 x changes of fresh PBS, followed by 5 minutes exposure to the ELF-97 fluorochrome according to the suppliers (Molecular Probes). Fluorescent micrographs were obtained using a DAPI long-pass filter to permit visualization of the novel yellow fluorescence from the UV wavelength excited fluorochrome.

10

Example 11: Monoclonal antibody production:

Dystrophin cDNA (cf27 in pUC plasmid, kindly provided by Prof. Kay Davies, University of Oxford) was digested with BamHI and NcoI. The 1640bp fragment from exon 4 to exon 16 was purified and ligated into pMW172 cut with the same restriction enzymes. After electroporation into E. coli BL21(DE3), protein expression was induced by 0.4 mM IPTG for 3 hours. Inclusion bodies were isolated by sonication and extracted sequentially with increasing concentrations of urea (2M, 4M, 6M and 8M in PBS). A 5 µg/ml solution of recombinant protein in 8M urea was used to immunize Balb/c mice and monoclonal antibodies were produced by the hybridoma method as described previously. Hybridoma supernatants were screened by elisa with recombinant proteins and positive wells (110 out of 288) were further tested for reaction with both native dystrophin (immunolocalization at muscle membrane) and denatured dystrophin (binding to 427kD band on Western blots of human muscle). Fourteen wells that passed this screening process were cloned twice by limiting dilution to establish the hybridoma lines. Ig subclass was determined using a mouse isotyping kit (Serotec Ltd.). Control blots with normal human lung showed that only one mAb (MANEX1011E) cross-reacted with utrophin.

25

Example 12: Epitope mapping:

Subconstructs of the pMW172:exon 4-16 construct were produced by PCR. Forward primers with added BamHI sites were synthesized by the UK Human Genome Mapping

30

Resource Center (Cambridge) as follows: exon 6 (ctcggatcccagggtcaaaaatgtaatg), exon 7 (ggggatccaggccagacctatttgac), exon 8 (ggggatccgatgttgataccacctatc), exon 10 (ggggatcccatttgaagctcctga) and exon 12 (ggggatcccatagagttttaatggatctc). The reverse primer in the pMW172 sequence was gttattgctcagcgggtggcagcag. PCR products were
5 digested with BamHI and EcoRI and cloned into pMW172 digested with the same enzymes. Each mAb was tested for binding to the expressed proteins on Western blots.

Example 13: Western Blotting

The method used for Western blotting of dystrophin was that of Arahata and
10 Hoffman. Cryomicrotome sections of 20 μ M thickness (10-20) from the various dog muscle samples were separately collected and stored at -80°C until gels were prepared for electrophoresis. Care was taken to be certain that fresh blades were used after positive control samples were sectioned. Electrophoresis was in 3.5-12% polyacrylamide at constant 60 volts for 16 hours. Proteins were electroblotted onto nitrocellulose (Amersham). The
15 same primary antibodies used above were used at 1:100 dilutions. Chemoluminescent detection was via a commercial kit used according to the direction provided by the supplier (ImmuneStar, goat anti-mouse, BioRad) and Kodak XR-1 film.

Example 14: Immunohistochemistry Frozen sections (6 μ M) of untreated triceps, injected
20 cranial tibialis, and normal cranial tibialis muscle were made using a Leica 3000 cryomicrotome and applied to Superfrost slides (Fischer Sci.). Primary monoclonal antibodies against dystrophin included a commercially available antibody specific for the carboxy-terminal region, (Novacastra) or exon 7 specific as described above. Primary antibody was applied directly to slides at 1:20 dilution in the presence of 5% normal goat
25 serum, while a goat anti-mouse secondary antibody labeled with FITC (Sigma or Jackson Immuesciences) was used to provide a fluorochrome for localization of dystrophin. Slides were counter-stained for 10 minutes with DAPI (Sigma) at 15 μ g/ml. Images were captured using 1/8 sec pixel accumulation as TIFF files with an Optronics cooled CCD camera and ScionImage frame-grabber installed in a PowerMac G3 and converted to Photoshop JPEG
30 files for printing on an HP 5M Color Laserprinter.

Example 15: Genomic PCR and Sequencing of exon 7

The protocol for genomic PCR of exon seven of the canine dystrophin gene was previously reported. PCR products were cloned into pCR-1 cloning vector (Invitrogen) and sequenced using an Applied BioSystems automated sequencer at the University of Miami

5 Cancer Center DNA Core.

Example 16: Chimeraplast used for GRMD mutation reversion

10 A diagram of the basis of the sequence of the chimeraplast is in Figure 1, Panel A. The chimeraplast is composed of a 5 base segment of DNA which defines the complement of the wild-type coding strand sequence at the splice acceptor site of intron six of the canine dystrophin gene [Sharp, N.H. et al. Genomics 13:115-121, 1992.] flanked by complementary segments of O-methyl-RNA (10-13 residues), two hairpin turns composed of 4 dT bases, a 3' 15 GC clamp segment, and a 5' complementary DNA strand which extends across either end of the two O-methyl-RNA segments. The native structure of such a molecule is believed to be a hairpin [Ye, S. Mol Med Today: 4:431-7, 1998.].

A timeline diagram of the experimental procedures performed on the GRMD affected male is found in Figure 1, Panel B. At six weeks of age (time point A), reverting 20 chimeraplast (200µg from BioSource) was mixed with 200 µg of calf thymus histone H1 (Sigma) and packaged in FuGENE™ 6 (Boehringer-Mannheim) plus Optimem (LTI) in a final volume of 5.0 ml and containing 7.5 µl/ml of fluorescent microspheres (Molecular Probes) to mark the site of injections. Using separate 100 µl injections, the full 5.0 ml was injected into the cranial tibialis compartment of the left limb and surgical biopsy samples 25 were taken and snap-frozen in liquid nitrogen at 2 (time point B) and 9 (time point C) weeks post-injection [Bartlett, R.J. et al. Nature Biology Short Reports. 9:163-164, 1998.] and at necropsy at 11 months (point E) post-injection from the left leg.

Without FuGENE™ 6 as the carrier vehicle

30

Additional chimera (from Kimeragen, Inc.) was injected into the contralateral limb during the surgical procedure for the 9 week biopsy (time point C), and at 6 weeks post-injection (time point D) for the contralateral limb, a biopsy was also taken from this limb. The protocol for injection was the same with the lone exception of deletion of the

- 5 FuGENE™ 6 from the injectate. Force generation studies (diagonal arrows) were performed at the three indicated times. The entire cranial tibialis, long digital extensor and triceps muscles (left and right) were removed at necropsy at 13 months of age when the animal was euthanized (time point E) due to progressive disease complications.

10 Example 17: RT/PCR of total RNA from canine skeletal muscle

- In panel A of Figure 2, the primers used in this study are positioned relative to the respective nucleotide sequence. The direction of the arrows indicate 5' (right pointing arrows) and 3' (left pointing arrows) primers. Frozen sections (20 µm) were extracted to isolate total RNA (RNAEasy, Promega) and RT/PCR was performed using 3' primers 704
15 and 120 and 5' primer 544 as previously published [Sharp et al. Genomics 13:115-121, 1992.]. In panel B, ethidium stained 1% agarose gels showing RT/PCR product from control and experimental tissues taken collected at the indicated times.

Example 18: In situ RT/PCR of canine muscle

- 20 See Figure 3. In brief, the method used for this analysis was extensive fixation of frozen section using 10% formalin overnight followed by brief pepsin treatment (17 minutes at 100 µg/ml) to permit infusion of RT/PCR reagents into the fixed tissue for reaction. Next, the tissue was treated with RNase-free DNase to remove nuclear DNA as template. Then using RT/PCR 3' primers from within exon 7 (459 and M23) and a 5' primer which spans
25 intron 6 (354) in genomic DNA (begins in exon 5 and ends in exon 6), RT/PCR was performed using the Roche/Boehringer Manneheim single tube Titan RT/PCR kit in the presence of dATP-Biotin to label all PCR product with biotin. Treatment with streptavidin conjugated with alkaline phosphatase (AP) and ELF-97 (Molecular Probes) fluorochrome were used to localize biotinylated PCR product. ELF-97 is a soluble, pale blue fluorescing
30 phosphate in original form, but upon cleavage by AP produces a precipitate that is brightly

yellow-green in fluorescence at the site of biotin incorporated into PCR product. The product is detected with a wide band-pass DAPI cube in a Leitz microscope. Arrows indicate localization of beads in panels F and H, and of specific RT/PCR product in experimental sections in panels E and G.

5

Example 19: Epitope mapping of dystrophin antibodies and western blotting of canine skeletal muscle

See Figure 4. In panel A, mixtures of recombinant proteins corresponding to exons 6-16 and 8-16 (upper blot) and exons 4-16, 7-16, 10-16 and 12-16 (lower blot) were loaded as a strip onto 12% acrylamide SDS-PAGE gels. After electroblotting, 14 mAbs were tested on each blot using a miniblottor apparatus as described previously (Thanh et al, 1995). The order of 13 mAbs shown is: MANEX6, 7B, 7C, 1011B, 1011C, 1011D, 1011A, 1011E, 1216E, 1216A, 1216B, 1216D, 1216C. MANEX7A is not shown but gave identical results to 7B and 7C. The relevant proteins are labeled and other protein bands on the blot are smaller degradation products of these. Note that MANEX1216E (lane 9) does not react with the smallest degradation product and hence recognizes a different epitope from 1216A-D; it is also the only MANEX1216 mAb to recognize native dystrophin in muscle sections (Table 1). In panel B, Western blotting lysed GRMD skeletal muscle proteins was performed according to a modified published protocol [Arahata, K. et al. Proc Natl Acad Sci USA 86:7154-8, 1989]. Frozen sections (20 μ m thickness, 10-20 total) were collected from untreated triceps muscle (lanes A and B), right cranial tibialis (CT) muscle (lane C), right long digital extensor (LDE) muscle (lane D), left LDE (lane E), CT muscle from a normal dog (lane F), and left CT muscle (lanes G-I). Sections were lysed in Lysis Buffer (1% SDS, 10 mM EDTA, Tris pH 8.0, and 50 mM DTT), boiled for 3 minutes, then cleared by centrifugation at 14,000 rpm in a microfuge for 5 minutes. Samples (3-10 μ l) were loaded onto 3.5-12% Laemmli gradient gels with 3% stacking gels and submitted to a constant voltage electric field of 60 volts per cm for 16 hours. Electroblotting was in transfer buffer (20% methanol, Tris glycine) for 3 hours in a Hoffer transblot chamber. An exon 7 specific antibody, MANEX7B (lane 2 in panel A, above), (1:100 dilution) in TBST was incubated for 60 minutes with the transferred membrane washed extensively, and probed with

chemoluminescent kit (ImmuneStar, BioRad) to detect the MANEX7B mAb bound to the membrane. Autoradiographic exposure of Kodak XL-R film was for 15 seconds. Samples were scanned using a UMAX Powerlook II scanner into Photoshop LE and archived on a UMAX Mac compatible computer.

5

Example 20: Immunofluorescence localization of dystrophin in canine muscle

See Figure 5. Frozen sections were blocked with normal goat serum, incubated with MANEX7B primary antibody and goat anti-mouse FITC conjugated secondary antibody and counter-stained with DAPI (15 µg/ml). Panels A-C were captured using the FITC
10 fluorescence bandpass filter while panels D-F were captured using a triple bandpass filter for DAPI fluorescence. Frames are shown for untreated GRMD triceps muscle (A and D), treated cranial tibialis muscle (B and E) and normal cranial tibialis muscle (C and F).

Example 21: Genomic PCR analysis of mutation reversion

15 See Figure 6. Genomic DNA was isolated from 20 µM frozen sections (20 each) from untreated triceps, treated cranial tibialis, and normal CT muscles using a commercial kit (Qiagen). PCR of genomic DNA was as previously reported using intronic primers that bracket exon 7 in the canine dystrophin gene [Bartlett, R.J. et al. Am J Vet Res: 57:650-4, 1996].

20 The mutant allele creates a novel Sau96 recognition sequence which was used to enrich for revertant alleles in all samples. After 10 cycles of PCR with the bracketing primers, all samples were digested with Sau96 to deplete GRMD alleles. Sau96 digested samples re-amplified using the same PCR conditions for 25 additional cycles. The 310 bp bands from each were separately ligated into the TA cloning vector pCR1 (Invitrogen) and
25 clones from each were submitted to analytical digestion with Sau96. All clones from the triceps muscle cut to completion indicative of the GRMD allele, and all clones from the normal CT muscle were resistant to digestion. Of 50 clones isolated from the treated CT muscle, 3 produced a pattern indistinguishable from the normal allele. Sequences of the normal (Top Panel), Sau96 resistant treated CT muscle (Middle Panel) and untreated triceps
30 muscle (Bottom Panel).

5

10

15

20

25

30

Claims

1. A composition for the modification of a mutant human dystrophin gene
5 comprising an oligonucleobase having both ribo-type and deoxyribotype nucleobases and
FuGENE™ 6, which oligonucleobase comprises:
 - a) a first and a second homologous region that are each at least 8 nucleobases
in length and together at least 20 and not more than 60 nucleobases in length, in which the
homologous regions are, respectively, homologous to a first fragment and a second fragment
10 of an exon of human dystrophin or of such exon and its 5' or 3' flanking intron, in which each
homologous region comprises at least 3 nucleobases of hybrid-duplex, and
 - b) a heterologous region that is disposed between the first and second
homologous region.
- 15 2. The composition of claim 1, in which each homologous region comprises a
segment of hybrid-duplex of at least 3 contiguous nucleobases.
3. The composition of claim 1, in which the ligand and oligonucleobase are
linked by a covalent linker.
20
4. The method of treating a subject dog or a human having muscular dystrophy
of the type that is treatable by the correction of a point mutation in the dystrophin gene of the
subject, which comprises:
providing a composition comprising an oligonucleobase and FuGENE™ 6, the
25 oligonucleobase having
 - a) a first and a second homologous region that are each at least 8 nucleobases in
length and together at least 20 and not more than 60 nucleobases in length, in
which the homologous regions are, respectively, homologous to a first
fragment and a second fragment of the dystrophin gene of the subject, which
30

- fragments are each adjacent to the point mutation, and in which each homologous region comprises at least 3 nucleobases of hybrid-duplex, and
- b) a heterologous region that is disposed between the first and second homologous region; and
- 5 administering to the subject an amount of the composition that is effective to ameliorate the subject's muscular dystrophy.

5. The method of claim 4, wherein the first and second fragment are fragments of an exon of the dystrophin gene or of such exon and the 3' or 5' flanking intron of the exon.

10

6. The method of claim 4, wherein the composition is administered to the subject by intra-muscular injection.

7. A method of treating a subject dog or human having muscular dystrophy of the type that is treatable by restoration of the reading-frame of the dystrophin gene of the subject, which comprises:

15

providing a composition comprising an oligonucleobase and FuGENE™ 6, the oligonucleobase having:

- a) a first and a second homologous region that are each at least 8 nucleobases in length and together at least 20 and not more than 60 nucleobases in length, in which the homologous regions are, respectively, homologous to a first fragment and a second fragment of the dystrophin gene of the subject, and in which each homologous region comprises at least 3 nucleobases of hybrid-duplex, and
- 20
- b) a heterologous region that is disposed between the first and second homologous region; and
- 25

administering to the subject an amount of the composition that is effective to ameliorate the subject's muscular dystrophy, such that an insertion or deletion that is adjacent to each fragment is introduced into the dystrophin gene of the subject.

30

8. The method of claim 7, wherein the first and second fragment are fragments of an exon of the dystrophin gene or of such exon and the 3' or 5' flanking intron of the exon.

9. The method of claim 7, wherein the composition is administered to the
5 subject by intra-muscular injection.

10

15

20

25

30

POINT MUTATION REVERSION IN THE DYSTOPHIN GENE

ABSTRACT

5 In the canine model of Duchenne muscular dystrophy in golden retrievers (GRMD), a point mutation within the splice acceptor site of intron 6 leads to deletion of exon 7 from the dystrophin mRNA and the consequent frameshift causes early termination of translation. A hairpin-shaped DNA and RNA chimeric oligonucleotide (chimeraplast or chimeraplast) has been designed to induce host cell mismatch repair mechanisms and revert the chromosomal
10 mutation to wild-type. Correction of this point mutation allows appropriate splicing of the dystrophin transcript to include exon 7. Direct skeletal muscle injection of the chimeraplast into the cranial tibialis (CT) compartment of a 6-week old affected male dog, and subsequent analysis of biopsy and necropsy samples, demonstrated an in vivo reversion of the GRMD mutation which was sustained for 11 months. RT/PCR analysis of exons 5-10 demonstrated
15 increasing levels of exon 7 inclusion with time. An exon 7-specific dystrophin antibody confirmed synthesis of normal-sized dystrophin product and positive localization to the sarcolemma. Chromosomal reversion in muscle tissue was confirmed by RFLP/PCR and sequencing the PCR product. This is the first long-term demonstration of reversion of a point mutation in muscle of a live animal using a chimeraplast. Before in vivo chimeraplasty
20 may become a viable alternative to myoblast transplantation or viral gene therapy for the treatment of Duchenne dystrophy and other muscular dystrophies, a viable systemic delivery method must be developed and tested in animals.

25

30

FIGURE 1

CHIMERAPLAST USED FOR GRMD MUTATION REVERSION

Figure 1A

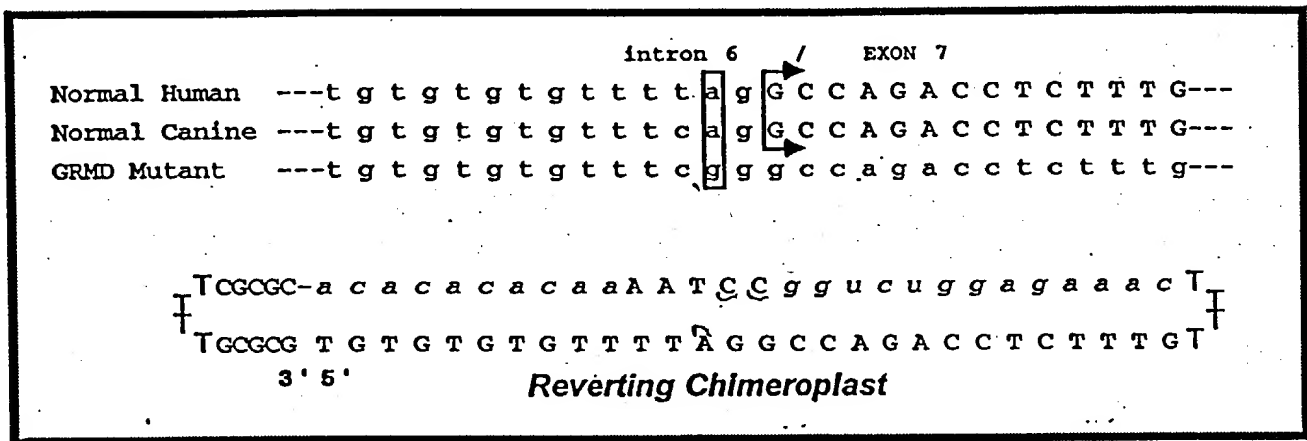
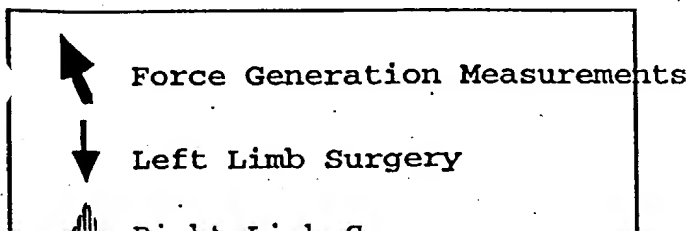
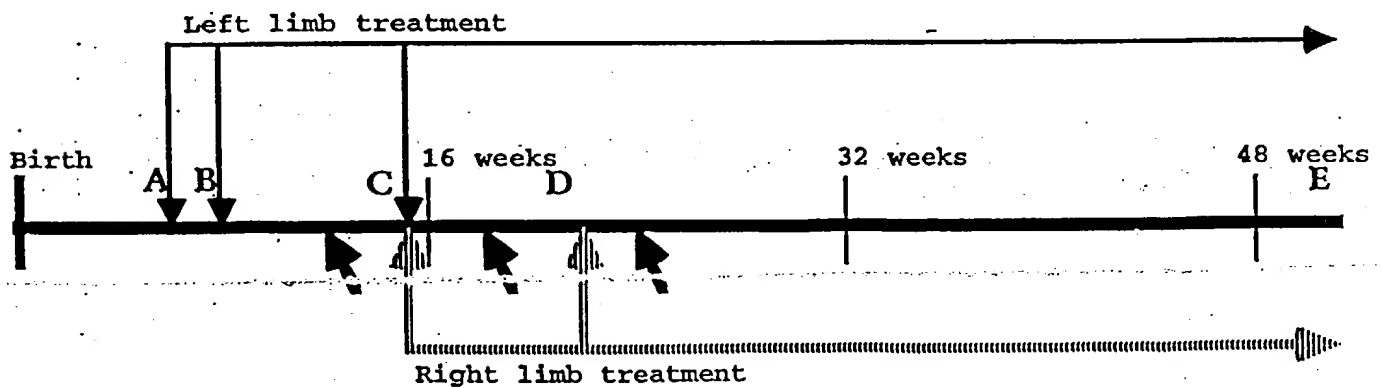


Figure 1B



RT/PCR OF TOTAL RNA FROM CANINE SKELETAL MUSCLE

The diagram illustrates the structure of the Dystrophin gene and its mRNA. The top part shows the Dystrophin cDNA, which is 14,000 bp long. It contains several exons, with Exon 7 being the focus. The cDNA sequence is shown with arrows indicating the direction of transcription (5' to 3'). Key positions are marked: 703, 544, 354, 120, 704, 121, and 122. The bottom part shows the genomic DNA structure of Exon 7, which is 735 bp long. It contains a mutation in the splice acceptor site, labeled M23, which causes a deletion of Exon 7 from the mRNA. The mutation is located at position 459. The diagram also shows the positions of exons 8 and 9, with a frame-shift from exon 8 to exon 9.

703 544 354

5' Dystrophin cDNA 3'

120 704 121 122

0 bp 1000 bp 2000 bp (14,000 bp)

Exon 7

355

735 bp 857 bp

M23 459

Mutation in Genomic DNA
in Splice Acceptor Site Causes
Deletion of Exon 7 from mRNA
(Frame-shift from exons 8 -> 9)

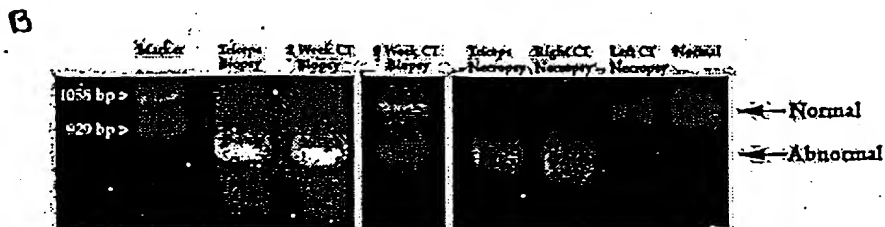


FIGURE 3

IN SITU RT/PCR OF CANINE MUSCLE

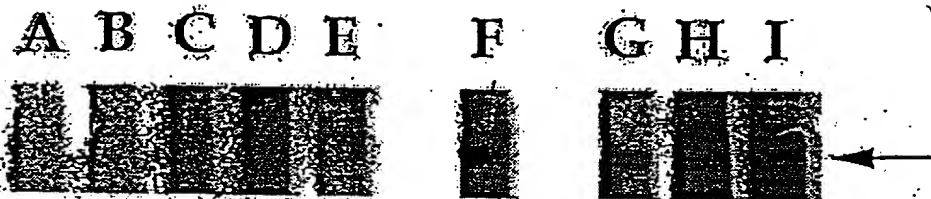


FIGURE 4

EPITOPE MAPPING OF DYSTROPHIN ANTIBODIES AND WESTERN BLOTTING OF
CANINE SKELETAL MUSCLE

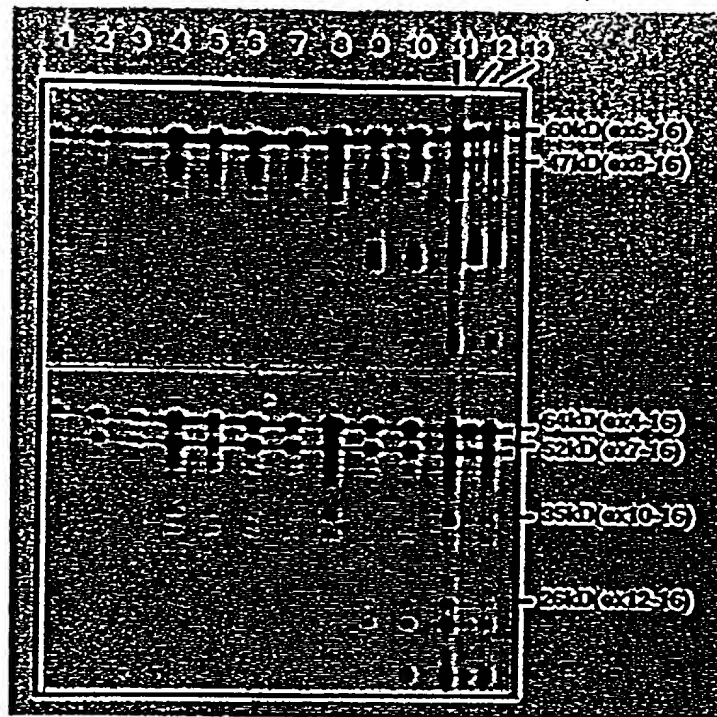


FIGURE 5

IMMUNOFLUORESCENCE LOCALIZATION OF DYSTROPHIN IN CANINE MUSCLE

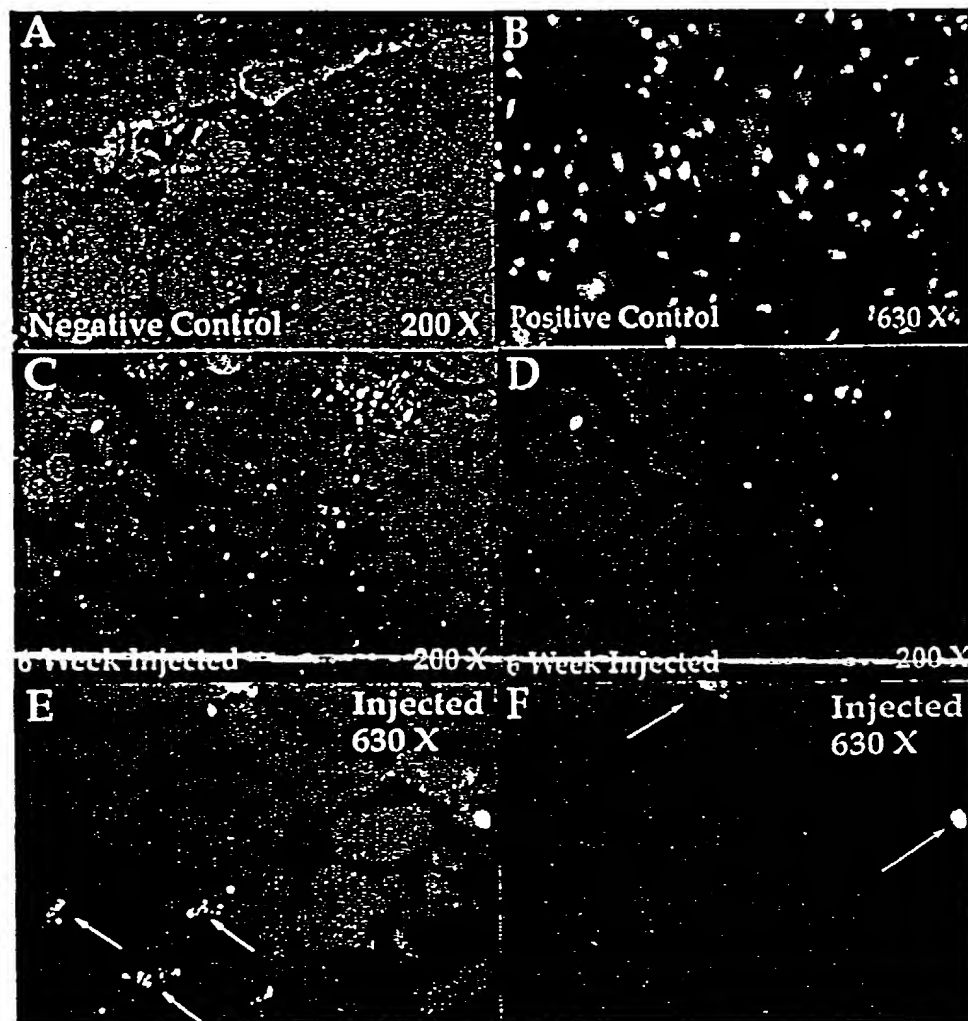


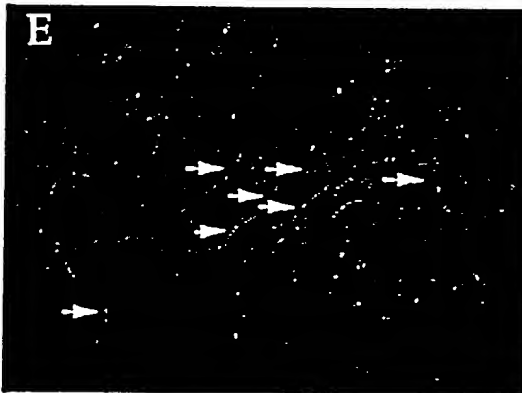
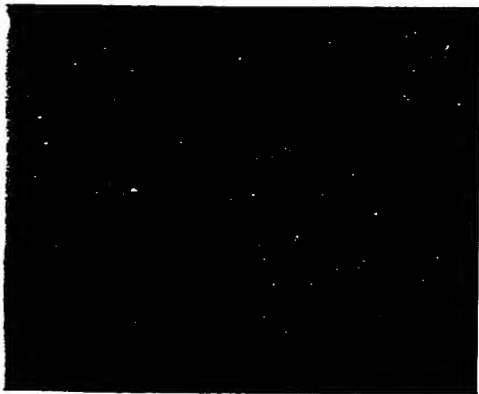
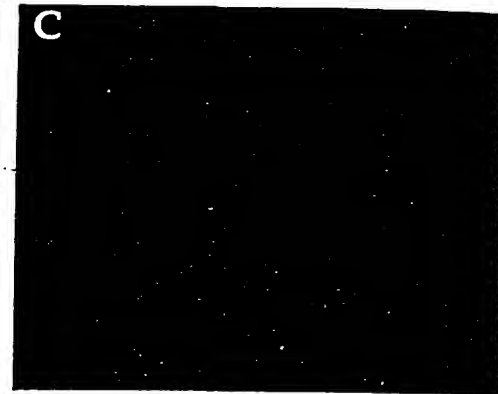
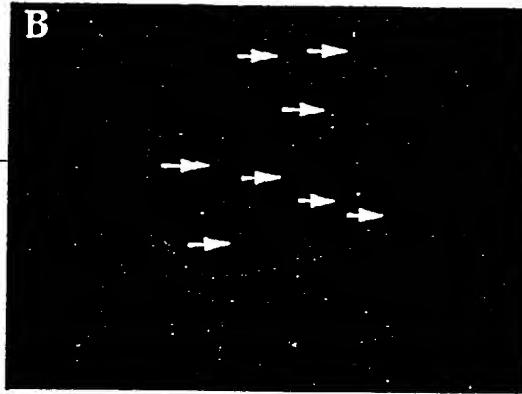
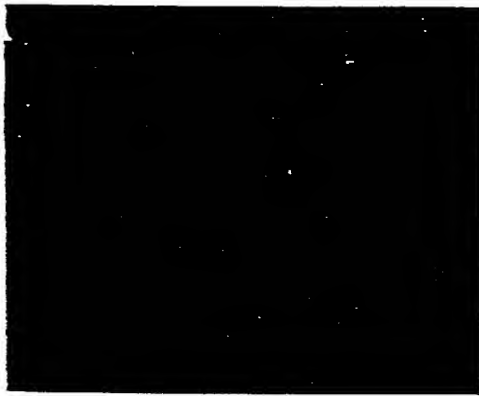
FIGURE 6

GENOMIC PCR ANALYSIS OF MUTATION REVERSION

Triceps (Negative)

CT (Experimental)

Normal (Positive)



6a

